

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2001, 15:50:48 ; Search time 24.95 Seconds  
(without alignments)  
1061.683 Million cell updates/sec

Title: US-08-978-174-1  
Perfect score: 1185  
Sequence: 1 MGPLRTVELFDVLSFYSW.....AHLGKWMGPIPPAVNARL 226

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL.15.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1185	100.0	226	4 Q9Y2Q3	Q9Y2Q3 homo sapien
2	1169	98.6	226	4 Q9P1S4	Q9P1S4 homo sapien
3	164.5	13.9	194	2 Q52782	Q52782 rhizobium 1
4	149	12.6	195	2 Q9X9Q7	Q9X9Q7 spingomona
5	145.5	12.3	312	5 Q22312	Q22312 caenorhabdi
6	140	11.8	197	2 Q85994	Q85994 spingomona
7	135	11.4	196	2 Q92HH4	Q92HH4 burkholderi
8	114	9.6	196	2 Q9WXH0	Q9WXH0 alcaligenes
9	106	8.9	199	2 Q923X5	Q923X5 pseudomonas
10	97.5	8.2	238	2 Q92167	Q92167 pseudomonas
11	91	7.7	581	2 Q9WXQ0	Q9WXQ0 thermocoga
12	90.5	7.6	410	1 Q26322	Q26322 methanobact
13	89	7.5	199	2 Q51499	Q51499 pseudomonas
14	89	7.5	592	10 Q9SPF1	Q9SPF1 arabidopsis
15	89	7.5	592	10 Q92SS6	Q92SS6 arabidopsis
16	83.5	7.0	308	2 Q92NM7	Q92NM7 exiguobacte
17	83.5	7.0	1007	10 Q92VD4	Q92VD4 arabidopsis
18	83	7.0	863	12 Q93124	Q93124 human calic
19	82.5	7.0	263	2 Q87197	Q87197 thermus aqu

20	82	6.9	203	2	007298	007298 pseudomonas
21	82	6.9	296	2	052799	052799 rhizobium 1
22	81	6.8	966	11	055098	055098 mus musculus
23	80.5	6.8	413	10	Q9LUP2	Q9LUP2 arabidopsis
24	80	6.8	221	5	009335	009335 nosema locu
25	80	6.8	402	1	026282	026282 methanobact
26	80	6.8	704	5	Q9VVT8	Q9VVT8 drosophila
27	80	6.8	821	2	Q30926	Q30926 escherichia
28	79	6.7	230	2	Q9RSB4	Q9RSB4 deinococcus
29	78.5	6.6	493	10	Q9SVJ3	Q9SVJ3 arabidopsis
30	78.5	6.6	693	5	Q9VS91	Q9VS91 drosophila
31	78.5	6.6	1072	6	077690	077690 bos taurus
32	78	6.6	248	1	029435	029435 archaeoglob
33	78	6.6	313	1	Q9YF18	Q9YF18 aeropyrum p
34	78	6.6	338	2	084200	084200 chlamydia t
35	78	6.6	704	2	056234	056234 escherichia
36	78	6.6	910	1	027560	027560 methanobact
37	78	6.6	1479	11	Q92277	Q92277 mus musculu
38	77.5	6.5	482	2	Q9XBN8	Q9XBN8 bacillus st
39	77.5	6.5	754	11	Q920R0	Q920R0 mus musculu
40	77.5	6.5	754	11	Q9JJJ4	Q9JJJ4 mus musculu
41	77.5	6.5	1166	4	Q9P2R3	Q9P2R3 homo sapien
42	77	6.5	253	1	Q90YA3	Q90YA3 pyrococcus
43	77	6.5	449	5	Q9V977	Q9V977 drosophila
44	77	6.5	659	2	Q928D0	Q928D0 chlamydia p
45	77	6.5	659	2	Q9JRY4	Q9JRY4 chlamydia p

## ALIGNMENTS

RESULT 1  
Q9Y2Q3 PRELIMINARY; PRT; 226 AA.  
ID Q9Y2Q3  
AC Q9Y2Q3; 1999 (TREMREL. 12, Created)  
DT 01-NOV-1999 (TREMREL. 12, Last sequence update)  
DT 01-NOV-1999 (TREMREL. 12, Last annotation update)  
DE GLUTATHIONE S-TRANSFERASE SUBUNIT 13 HOMOLOG.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mao M., Ye M., Zhang Q., Zhou J., Wu J., Shen Y., Kan L., He K.,  
RA Gu B., Fu G., Chen S., Chen Z.;  
RL "Human GSTK1-1 homolog gene."  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF070657; AAD20963.1;  
KW Transferase.  
SQ SEQUENCE 226 AA; 25497 MW; D3FDAFDI533B58A4 CRC64;

Query Match	100.0%	Score 1185;	DB 4;	Length 226;
Best Local Similarity	100.0%	Pred. No. 4.9e-96;		
Matches 226;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy 1	MGPLRTVELFDVLSFYSWGLGFEILCRYNQINWINLQRLPSLIITGIMKDSGNKPPGLLP	60		
Db 1	MGPLRTVELFDVLSFYSWGLGFEILCRYNQINWINLQRLPSLIITGIMKDSGNKPPGLLP	60		
Qy 61	RKGLYMANDLKLLRHHLQIPHFPKDFLSVLMKGLSASAMRFLTAVNLEHPHMELEKASRE	120		
Db 61	RKGLYMANDLKLLRHHLQIPHFPKDFLSVLMKGLSASAMRFLTAVNLEHPHMELEKASRE	120		
Qy 121	LMRYVMSRNEDITEPOSILAAAEKAGMSAEOAGLLEKIATPKVKNQLKETTAAACRYGA	180		
Db 121	LMRYVMSRNEDITEPOSILAAAEKAGMSAEOAGLLEKIATPKVKNQLKETTAAACRYGA	180		
Qy 181	FGLPITVAHVGDGTHMLFGSDRMELLALHLLGKWMGPPIPPAVNARL	226		
Db 181	FGLPITVAHVGDGTHMLFGSDRMELLALHLLGKWMGPPIPPAVNARL	226		

•

10

---

\_\_\_\_\_

NCBI\_TaxID=6239;  
 [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favellio A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirstern J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
 RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;  
 RL Nature 0:0-0(0).  
 [2]  
 SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Favellio A.;  
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
 [3]  
 SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U13643; AAA21082.1; -;  
 SQ SEQUENCE 312 AA; 35436 MW; 0FB8561F57687120 CRC64;

Query Match 12.3%; Score 145.5; DB 5; Length 312;  
 Best Local Similarity 22.6%; Pred. No. 6.3e-05;  
 Matches 51; Conservative 46; Mismatches 100; Indels 29; Gaps 6;  
 QY 1 MGPLPRTVELFYDVLSPYSW-----LGFELICRYQ-----NIWNINLQ 38  
 DB 1 MAPLPR-VKCYFDVCPNSWITITQATSHNSLFRIDFPVCDKIGILHNAQIWNRRQ 59  
 QY 39 LRPSLI--TGIMKDSGNKPPGLPRKGLYMAND---LKLRLHHLQIPHFPRKDELSVMLE 93  
 DB 60 VHSNLSWTKTIEVPEQAESEETISENGILOKIDERGKLLICERVVPVDPKNTYKTAVA 119  
 QY 94 KGSLSAMRFLTAVNLEHPEMLEKASRELWVRVSRNEDITEPQSILAAAEKAGMSAEQAO 153  
 DB 120 RGSVPQLFLTSTIREQYDOLYRIHHLKRLQRLPVHYGCHMSTVYCRELGISFKIAE 179  
 QY 154 GLEKIATPKVNQKLETTAEACRYGAFGLPITVAHYD-GOTHMLF 198  
 DB 180 DIVARLSSPENRSILHKNCKEAVDFKLTAPGLILLTDEGDTIKIF 225

RESULT 6  
 O85994 PRELIMINARY; PRT; 197 AA.  
 ID O85994;  
 AC O85994;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE 2-HYDROXYCHROMENE-2-CARBOXYLATE ISOMERASE.  
 GN NAHD.  
 OS Sphingomonas aromaticivorans.  
 OC Plasmid pNL1.  
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;  
 OC Sphingomonas.  
 QX NCBI\_TaxID=48935;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-F199;  
 RA Romine M.F., Stillwell L.C., Wong K.-K., Thurston S.J., Sisk E.C.,  
 RA Sensen C.W., Gaasterland T., Saffer J.D., Fredrickson J.K.;  
 RT "Complete sequence of a 184 kb catabolic plasmid from Sphingomonas  
 RT aromaticivorans strain F199."  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF079317; AAD04010.1; -;

KW Isomerase; Plasmid.  
 SQ SEQUENCE 197 AA; 21998 MW; COCD74E7106C07CE CRC64;

Query Match 11.8%; Score 140; DB 2; Length 197;  
 Best Local Similarity 23.1%; Pred. No. 0.0001;  
 Matches 50; Conservative 40; Mismatches 94; Indels 32; Gaps 7;  
 QY 4 LPRTVELFYDVLSPYSWGLFEILCRYQNIWNINLQRLPSLTITGIMKDSGNKPPG---LPL 50  
 DB 1 MTRIDFDFDISPISYLAQLKLPETARAAGCTVDYPIPEAKIRAGNYGSPNREVLP 60  
 QY 61 RKGLYMANDKLRLHHLQIPHFPKDF-----LSVMLEKGSLSAMRFLTAVNLEHPEML 114  
 DB 61 -KIRVMKADLERWAERYGVPLTFPASAFACADWNCVLFAREHGAFAEVT-----109  
 QY 115 EKASRELWVRVSRNEDITEPQSILAAAEKAGMSAEQAOGLLEKIATPKVNQKLETTAE 174  
 DB 110 -----DAYRRIWGQIDPGDRNELAAACATAAGL---DPAALIAFVESPAGONEYKARSQ 161  
 QY 175 ACRYGAFGLPITVAHYDVGQTHMLFGSDRMELLHL 210  
 DB 162 AIQGVYGAFL--MFVDDQ--IFWGNDRDLFLAEYL 193

RESULT 7  
 Q92HH4 PRELIMINARY; PRT; 196 AA.  
 ID Q92HH4;  
 AC Q92HH4;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE ISOMERASE PHND.  
 GN PHND.  
 OS Burkholderia sp. RP007.  
 OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
 OC Burkholderia.  
 QX NCBI\_TaxID=83784;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RP007.  
 RX MEDLINE-99102215; PubMed-9882667;  
 RA Laurie A.D., Lloyd-Jones G.;  
 RT "The phn genes of Burkholderia sp. strain RP007 constitute a divergent  
 RT gene cluster for polycyclic aromatic hydrocarbon catabolism."  
 RL J. Bacteriol. 181:531-540(1999).  
 DR EMBL: AF061751; AAD09871.1; -;  
 KW Isomerase.  
 SQ SEQUENCE 196 AA; 21908 MW; 58716E9039BEA768 CRC64;

Query Match 11.4%; Score 135; DB 2; Length 196;  
 Best Local Similarity 24.0%; Pred. No. 0.00028;  
 Matches 50; Conservative 35; Mismatches 103; Indels 20; Gaps 5;  
 QY 7 TVELEYDVLSPYSWGLFEILCRYQNIWNINLQRLPSLTITGIMKDSGNKPPGL--LPRKGL 64  
 DB 2 TIDFFDFLSPYAYLARHRLTQVALHGCALAYKPIDLARAKLAIGNTGPNRDMVPKLA 61  
 QY 65 YMANDKLRLHHLQIPHFPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRELWVR 124  
 DB 62 YVVEDLKRWAARYRIPIEFIKNTKRMVGT-----FYAEARGOQADYVQRAYHLAWGE 116  
 QY 125 VWSRNEDITEPQSILAAAEKAGMS-ABQAOGLLEKIATPKVNQKLETTAEACRYGAFGL 183  
 DB 117 GGAPDD-----AALRSIAVSMGMDAADFLFLDSSEAEATYNSTLEAISAGVGV 168  
 QY 184 PITVAHYDVGQTHMLFGSDRMELLHL 211  
 DB 169 PTMAYGRD-----MWWGNDRIDFLFLEHLG 192

RESULT 8

```

KW Isomerase; Plasmid.
SQ SEQUENCE 199 AA; 23067 MW; 219466E06AC04E47 CRC64;

Query Match      8.9%; Score 106; DB 2; Length 199;
Best Local Similarity 24.9%; Pred. No. 0.099; 91; Indels 34; Gaps
Matches 52; Conservative 32; Mismatches 91; Indels 34; Gaps 34;

QY 8 VELYDVLSPYSWLGFELICRYQNIWNINQLRPSLTIGIMKDSGNKPPGL--LPKRGLY 65
   | : : | | | : | | : : : | | : | | : | | : | | : | | : | | : |
Db 5 VDFYDFLSPISYLANHRLSVLAGRYGFSIQYHAIDLARAKTAIGNIGPSNRDLKVKLDY 64
   | : : | | | : | | : : : | | : | | : | | : | | : | | : | | : |
QY 66 MANDLKLRLHUIQIPHFPEKDFLSVMLEKGSLSAMRFLTAVNLEHPMELEKASR----- 119
   | : : | | | : | | : : : | | : | | : | | : | | : | | : | | : |
Db 65 LKVDIQWADLYRIPLEFPNFNSRRYNAG-----LYTPAARERAAEYVRLVF 112
   | : : | | | : | | : : : | | : | | : | | : | | : | | : | | : |
QY 120 -ELMNRVSRNRNEDITEPQSILA-AAEKAGMSAEQAQGLLEKIATPKVKQLKETTAAACR 177
   | : : | | | : | | : : : | | : | | : | | : | | : | | : | | : |
Db 113 DSAWKGWALDAD-----SLLAEVCDKLNWDLGFEEDFLNSENAAYD--BETQRAID 164
   | : : | | | : | | : : : | | : | | : | | : | | : | | : | | : |
QY 178 YGAFGLPTVAHVHGQTHMLFGSDRMELL 206
   | | : | | | | | : | | : | | : | | : | | : | | : | | : | | : |
Db 165 RKVFCGVP-TVFWDD---QMWGNDRLFML 189
   | | : | | | | | : | | : | | : | | : | | : | | : | | : | | : |

RESULT 10
Q92167 PRELIMINARY; PRT; 238 AA.
ID Q92167 AC
AC Q92167;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE 2-HYDROXYCHROMENE-2-CARBOXYLATE DEHYDROGENASE.
GN NAHD.
OS Pseudomonas stutzeri (Pseudomonas perfectomarina).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=316;
OX {1}
RN SEQUENCE FROM N.A.
RP STRAIN=AN10.
RC Bosch R., Lalucat J., Timmis K.N., Moore E.R.B.;
RA "Complete nucleotide sequence of a chromosomally encoded naphthalene
RT degradation pathway from Pseudomonas stutzeri AN10 and its
RT evolutionary significance."
RL Submitted (Dec-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF039533; AAD02142.1; ..
SQ SEQUENCE 238 AA; 26359 MW; 29B972C1CCF7C71B CRC64;

Query Match      8.2%; Score 97.5; DB 2; Length 238;
Best Local Similarity 23.5%; Pred. No. 0.69;
Matches 53; Conservative 37; Mismatches 103; Indels 33; Gaps

QY 8 VELYDVLSPYSWLGFELICRYQNIWNINQLRPSLTIGIMKDSGNKPPGL--LPKRGLY 65
   | : : | | | : | | : : : | | : | | : | | : | | : | | : | | : |
Db 19 VDFYDFLSPISYLANHRLSVLAGRYGFSIQYHAIDLARAKTAIGNIGPSNRDLKVKLAY 78
   | : : | | | : | | : : : | | : | | : | | : | | : | | : | | : |
QY 66 MANDLKLRLHUIQIPHFPEKDFLSVMLEKGSLSAMRFLTAVNLEHPMELEKASRELWMR 124
   | : : | | | : | | : : : | | : | | : | | : | | : | | : | | : |
Db 79 LMVDLKRWAELYGLFFLPFANNYSQRMNAGLYSGAETQTAAYV-----NTVFNA 128
   | : : | | | : | | : : : | | : | | : | | : | | : | | : | | : |
QY 125 VWSNRNEDIT-EPQSILAAAEKAGMSAEQAQGLLEKIATPKVKQLKETTAAACRYAGFL 183
   | : : | | | : | | : : : | | : | | : | | : | | : | | : | | : |
Db 129 VW--GEGIALDSESLALV--CGTLGWDAFAAEFFLSDAATNAYDEHTQAAIERKVFGV 184
   | : : | | | : | | : : : | | : | | : | | : | | : | | : | | : |
QY 184 "PIITVAHVHGQTHMLFGSDRMELLALL-----GEKMWGPI 218
   | : : | | | : | | : : : | | : | | : | | : | | : | | : | | : |
Db 185 P-TMFLGD---QMWGNDRLFMLENTLRCCSGAGIAAAGETGVKVP 226
   | : : | | | : | | : : : | | : | | : | | : | | : | | : | | : |

RESULT 11
R9WX00

```

```

ID Q9WXQ0 PRELIMINARY; PRT; 581 AA.
AC Q9WXQ0;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE ABC TRANSPORTER, ATP-BINDING PROTEIN.
GN TM0043.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima."
DR EMBL: AE001691; AAD35137.1; -
DR HSSP: P13569; INBD.
DR TIGR: TM0043; -
DR INTERPRO: IPR001140; -
DR INTERPRO: IPR001617; -
DR INTERPRO: IPR002106; -
DR PFAM: PF00005; ABC_tran; 1.
DR PFAM: PF00664; ABC_membrane; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
DR PROSITE: PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
KW ATP-binding.
SQ SEQUENCE 581 AA; 66324 MW; 55719E0DCA5D0A05 CRC64;

Query Match 7.7%; Score 91; DB 2; Length 581;
Best Local Similarity 19.0%; Pred. NO. 8.4; Mismatches 67; Indels 38; Gaps 6;
Matches 35; Conservative 44;

QY 28 RYQNIW-----NINLQRLPSLITGIMKDSGNKPPGL-----IPRKGLYMAN 68
DB 343 RFENVWESYDGNKWLKIDNLFQPKLYAIVGTGGKSTLSLNGLYIPQKGNIFD 402
QY 69 DLKLRHHLQIP----IHFPKDFL-----SVMLEKGLSAMRFLTAVNLEHP-EMLE 115
DB 403 EIPLLYNLKLVRKQIAAYPQDVLFLSGTLDNIRLFDSEIPERVLKAKRVHLDIIE 462
QY 116 KASRELWVRWNRNEDITEPQSTLAAAEKAGM-----SAPQAGLLEKIATPKVKNOLK 169
DB 463 RLPQGVYVEIVERGTTLSAGERQLIARAVLFDKIFILDEATSNVDVITKIQEAL 522
QY 170 ETTE 173
DB 523 ELSK 526

RESULT 12
ID O26322 PRELIMINARY; PRT; 410 AA.
AC O26322;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE FLAVOPROTEIN A HOMOLOG (II).
GN MTH220.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanobacterium.
OX NCBI_TaxID=2166;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=DELTA H.;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hwang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nollung J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics."
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL: AE000809; AAB84726.1; -
DR INTERPRO: IPR001279; -
DR PFAM: PF00753; lactamase_B; 1.
SQ SEQUENCE 410 AA; 45740 MW; 8EAF9D3B363A8BD5 CRC64;

Query Match 7.6%; Score 90.5; DB 1; Length 410;
Best Local Similarity 25.1%; Pred. NO. 5.8;
Matches 65; Conservative 31; Mismatches 82; Indels 81; Gaps 16;

QY 13 DVLSPSYSLGFEILC-----RYQNIWNLQRLPSLITGIMKDSGNK----- 54
DB 99 EVLKRY---GSEIITAKAEGRLROHYSIPQDTPMQ---TVKTGDDTLGKTLTFLEAP 152
QY 55 ----PPG---LLPRKGLYMANDKLLRHHLQIPHFPPKDFLSVMLEKGLSAMRF-----L 103
DB 153 MLHWPDMSMFLLEEGILFSND--AFGOHLCKISKREDKOVPEAVLMD---AAMKFYANLL 207
QY 104 TAVNLEHPMLEKAS--RELWM-----RVWSRNEDETEPOSILAAAEKAGMSA 149
DB 208 TPLS---PLVLRKFESEVKELGELKIGMTAPSHGQIW-----TEPLKIIAA-----YT 252
QY 150 EQAOGILLEKIATPKVKNOLKETTAACRYGAGFLPITVAHVHVGDTGTHLFGSDRMELLHL 209
DB 253 DWATGKCRDKAT-IYDTMHYSTRMLAHAMAEGEL--MAADVDSNHHFHEDESEIVKNI 309
QY 210 LGER-----WVGPIIP 219
DB 310 LESKAVFTGSPFMFGPPF 328

RESULT 13
ID Q51499 PRELIMINARY; PRT; 199 AA.
AC Q51499;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)
DE ISOMERASE.
GN PAHE.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAK1;
RA Takizawa N., Iida T., Yamauchi K., Satoh S., Wang Y., Fukuda M.,
RA Kiyohara H.;
RT "The molecular analysis of an NAH7-type gene cluster, pah, located on
RT the chromosome of Pseudomonas aeruginosa pak1."
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: D84146; BAA12247.1; -
KW Isomerase.
SQ SEQUENCE 199 AA; 22509 MW; 81C2A90CA56E5462 CRC64;

Query Match 7.5%; Score 89; DB 2; Length 199;
Best Local Similarity 24.2%; Pred. NO. 3;
Matches 51; Conservative 37; Mismatches 95; Indels 28; Gaps 9;

```



Search completed: February 15, 2001, 15:53:11  
Job time: 143 sec

---





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2001, 15:50:42 ; Search time 16.3 Seconds  
(without alignments)  
474.098 Million cell updates/sec

Title: US-08-978-174-1

Perfect score: 1185

Sequence: 1 MGPLPRTVELFDVLSFYSW.....AHLGKWMGPIPPAVNARL.226

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_36.\*

- 1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1185	100.0	226	Y77499	Novel human glutat
2	1185	100.0	256	Y59988	Human endometrium
3	89	7.5	545	Y32947	Mutant threonine d
4	89	7.5	545	Y05711	Feedback insensiti
5	89	7.5	590	Y32950	Mutant threonine d
6	89	7.5	592	Y32939	Mutant threonine d
7	89	7.5	592	Y32948	Mutant threonine d
8	89	7.5	592	Y32951	Wild type threonin
9	89	7.5	592	Y05702	Arabidopsis wild-t
10	89	7.5	592	Y05703	Feedback insensiti
11	89	7.5	600	Y32952	Mutant threonine d
12	89	7.5	609	Y32940	Mutant threonine d

13	89	7.5	609	20	Y05704	Feedback insensiti
14	83.5	7.0	308	17	W04266	Inosine-guanosine
15	83.5	7.0	539	20	Y32942	Mutant threonine d
16	83.5	7.0	539	20	Y05706	Feedback insensiti
17	81	6.8	911	20	Y59957	Mouse STE20-relate
18	80.5	6.8	339	20	Y33825	Amino acid sequenc
19	80.5	6.8	532	20	Y32943	Mutant threonine d
20	80.5	6.8	532	20	Y05707	Feedback insensiti
21	79.5	6.7	1073	18	W32063	Human ST receptor
22	79.5	6.7	1073	19	W37371	Human ST receptor
23	78	6.6	341	20	Y37006	Amino acid sequenc
24	77	6.5	659	20	Y35014	Chlamydia pneumoni
25	76.5	6.5	216	18	W34203	Streptomyces non-m
26	76.5	6.5	216	19	W55804	Streptomyces rosec
27	76.5	6.5	317	20	Y39338	Carboxymethyl cell
28	76.5	6.5	317	20	Y23764	A carboxymethyl ce
29	76.5	6.5	317	21	Y56814	T. maritima thermo
30	76.5	6.5	497	13	R27786	Adrenodoxin reduct
31	76.5	6.5	497	16	R66893	Human adrenodoxin
32	76	6.4	588	20	Y30124	Human adrenodoxin
33	75.5	6.4	1398	17	R87008	A human protein wi
34	75.5	6.4	1398	18	W24124	Protease. Pyococc
35	75.5	6.4	1398	20	W94839	Pyrococcus furiosu
36	74	6.2	342	21	Y91567	WO9856926 Seq ID 6
37	74	6.2	492	10	P94363	Human secreted pro
38	74	6.2	554	21	Y91569	Rovine adrenal gla
39	74	6.2	912	20	Y59399	Human secreted pro
40	74	6.2	968	20	Y59366	Human GSK2 protein
41	73.5	6.2	869	15	R56501	Full length huma
42	73.5	6.2	869	15	R56501	TATA-binding prote
43	73.5	6.2	869	17	W06091	Human TATA-binding
44	73.5	6.2	869	18	W25022	TATA-binding prote
45	73.5	6.2	985	20	W30811	Arabidopsis thalia
					W30807	Arabidopsis thalia

## ALIGNMENTS

RESULT 1

Y77499  
ID Y77499.standard; Protein; 226 AA.

AC Y77499;

DT 05-JUN-2000 (first entry)

DE Novel human glutathione S-transferase, GSTS.

KW Glutathione S-transferase; human; GSTS; cancer; immune disorder;  
KW gene therapy; diagnosis; treatment; drug screening.

OS Homo sapiens.

PN US6030809-A.

PD 29-FEB-2000.

PE 25-NOV-1997; 97US-0978174.

PF 25-NOV-1997; 97US-0978174.

PR (INCYTE), INCYTE PHARM INC.

PI Hillmap JL, Shah P, Lal P, Corley NC;

XX WPI; 2000-205204/18.

DR N-PSDB; 202599.

PT Isolated nucleic acid encoding glutathione S-transferase useful in the

PT production of agents for preventing, diagnosing and treating diseases

PT associated with cell proliferation -

XX Claim 8; Fig 1A-C; 27pp; English.

100

XX 19-AUG-1999.  
 XX PD  
 XX PF  
 XX PF 08-JAN-1999; 98WO-US00560.  
 XX XX  
 XX PR 10-JUL-1998; 98WO-US14362.  
 XX PR 17-FEB-1998; 98US-0074875.  
 XX XX  
 XX PA (DOWC ) DOW AGROSCIENCES LLC.  
 XX PA (PURD ) PURDUE RES FOUND.  
 XX XX  
 XX PI Larrinua IM, Merlo DJ, Mourad GS, Paredy DR;  
 XX XX WPI; 1999-527375/44.  
 XX DR N-PSDB; 211205.  
 XX XX  
 XX PT New nucleic acid encoding threonine dehydratase deaminase resistant  
 XX PT to feedback inhibition, useful as selection marker for cell  
 XX PT transformation and to impart herbicide resistance  
 XX PS Claim 13; Page 131-134; 194pp; English.  
 XX CC This sequence represents a mutant Arabidopsis thaliana threonine  
 XX CC dehydratase/deaminase (TD) protein of the invention. The protein is a  
 XX CC feedback insensitive mutant. The TD DNA sequence is used as molecular  
 XX CC marker (imparting resistance to toxic structural analogues of isoleucine)  
 XX CC for selecting transformed cells and to produce transformants with  
 XX CC increased levels of isoleucine (and thus better nutritional value) or of  
 XX CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for  
 XX CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also  
 XX CC TD-expressing plants permit use of the isoleucine structural analogues as  
 XX CC herbicides. The DNA sequences are alternatives for antibiotic resistance  
 XX CC markers (which are potentially harmful to the environment). Since no  
 XX CC human analog of TD exists (humans can not synthesize isoleucine), it  
 XX CC should be safe to use.  
 XX XX  
 XX SQ Sequence 545 AA;

Query Match 7.5%; Score 89; DB 20; Length 545;

Best Local Similarity 22.8%; Pred. No. 0.3;  
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;

QY 55 PPGL-LPRKGLYMANDLKLRLHLLQIPHFDPKDFLSVM-----LEKGSLS-AMRFLT-- 104  
 DB 5 PPKLPIR-----LKVSPNSIQ-----YPAGYLGAVPERTNEAENGSAEAMEYLTNI 52  
 QY 105 -----AVNLEHP-EMLEKASRELMMRWVSRNEDITE-----POSTLAA 141  
 DB 53 LSTKVYDIAIESPLQLAKKLSKRGVRYMLKREDIQPVFSFKLRGAYNMVKKIPADGLAK 112  
 QY 142 AEKAGMSAEQAQG-----LLEKIATPKVKNQKLTETEAACRYGAGLPTIVA 188  
 DB 113 GVCSAGNHAGQVALSASKIGCTAVIMPTTPEIKQAVENL----- 156  
 QY 189 HVDGQTHMFGSDRMELLAHL---LGEKWMGPIPP 220  
 DB 157 ---GATVVLFGSDYDQAQAHAKIRAEEGLTFIPP 188

RESULT 4

Y05711 standard; Protein; 545 AA.

XX Y05711;

AC Y05711;

XX 19-JUL-1999 (first entry)

DT Feedback insensitive threonine dehydratase/deaminase.

DE Threonine dehydratase/deaminase; OMRI gene; feedback inhibition;

XX transgenic plant; selectable marker; isoleucine; mutant.

KW

XX

OS Arabidopsis thaliana.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 XX PH 439..457  
 XX FT /note= "regulatory region R4"  
 XX FT 489..507  
 XX FT /note= "regulatory region R6"  
 XX FT Misc-difference 452  
 XX FT /note= "Arg in wild-type TD"  
 XX FT Misc-difference 497  
 XX FT /note= "Arg in wild-type TD"

XX WO9902656-A1.

XX 21-JAN-1999.

XX 10-JUL-1998; 98WO-US14362.

XX 17-FEB-1998; 98US-0074875.

XX 10-JUL-1997; 97US-0052096.

XX (PURD ) PURDUE RES FOUND.

XX Mourad GS;

XX WPI; 1999-120860/10.

XX N-PSDB; X25340.

XX New sequences encode mutant threonine dehydratase/deaminase - which

XX is insensitive to feedback inhibition, useful as a selective marker

XX to produce transformed cells resistant to toxic isoleucine analogues

XX Disclosure; Page 71-73; 120pp; English.

XX The present sequence represents an Arabidopsis thaliana mutant

XX threonine dehydratase/deaminase (TD) protein which, unlike

XX wild-type TD, is insensitive to feedback inhibition by isoleucine.

XX Claimed polynucleotides (see X25332-40), originally isolated and

XX cloned from A. thaliana mutated line GM11b (Omrl/Omrl), encode

XX feedback insensitive TD that can be used to transform a wide

XX variety of plants, fungi, bacteria and yeast. Mutant TD differs

XX from the wild-type enzyme only by an R499C amino acid substitution

XX regulatory region R4, and by an R544H substitution in regulatory in

XX region R6. Mutant TD is not only insensitive to feedback

XX inhibition by isoleucine, but is also insensitive to structural

XX analogues of isoleucine that are toxic to plants and microorganisms

XX which synthesize only wild-type TD. Nucleotide sequences encoding

XX mutated forms of TD can therefore be used to create cells that are

XX insensitive to compounds normally toxic to cells expressing only

XX wild-type TD enzymes, and thus may be used to provide a biochemical

XX selectable marker. Transformants that express the mutant TD show

XX increased levels of isoleucine production, and thus provide an

XX improved nutrient source.

XX Sequence 545 AA;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

QY 189 HVDGTHMLFGSDRMELLAHL---LGEKWMGPIPP 220  
 ID Y32950 standard; Protein; 590 AA.  
 AC Y32950;  
 XX  
 DT 09-NOV-1999 (first entry)  
 DE Mutant threonine dehydratase/deaminase protein sequence.  
 DE Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; mutin.  
 XX Arabidopsis thaliana.  
 OS Synthetic.  
 XX WO9941395-A1.  
 PN 19-AUG-1999.  
 PD 08-JAN-1999; 99WO-US00560.  
 PE 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 PA (PURD ) PURDUE RES FOUND.  
 XX Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;  
 XX WPI; 1999-527375/44.  
 DR N-PSDB; Z11208.  
 XX New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX Example 3; Page 119-123; 194pp; English.

QY 189 HVDGTHMLFGSDRMELLAHL---LGEKWMGPIPP 220  
 ID Y32950 standard; Protein; 590 AA.  
 AC Y32950;  
 XX  
 DT 09-NOV-1999 (first entry)  
 DE Mutant threonine dehydratase/deaminase protein sequence.  
 DE Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; mutin.  
 XX Arabidopsis thaliana.  
 OS Synthetic.  
 XX WO9941395-A1.  
 PN 19-AUG-1999.  
 PD 08-JAN-1999; 99WO-US00560.  
 PE 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 PA (PURD ) PURDUE RES FOUND.  
 XX Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;  
 XX WPI; 1999-527375/44.  
 DR N-PSDB; Z11208.  
 XX New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX Example 3; Page 119-123; 194pp; English.

This sequence represents a mutant Arabidopsis thaliana threonine dehydratase/deaminase (TD) protein of the invention. The protein is a feedback insensitive mutant. The TD DNA sequence is used as molecular marker (imparting resistance to toxic structural analogues of isoleucine) for selecting transformed cells and to produce transformants with increased levels of isoleucine (and thus better nutritional value) or of intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also TD-expressing plants permit use of the isoleucine structural analogues as herbicides. The DNA sequences are alternatives for antibiotic resistance markers (which are potentially harmful to the environment). Since no human analog of TD exists (humans can not synthesize isoleucine), it should be safe to use.

Sequence 590 AA;

Query Match 7.5%; Score 89; DB 20; Length 590;  
 Best Local Similarity 22.8%; Pred. No. 0.34;  
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;

QY 55 PPGL-LPRKGLYMANDLLRHLHLPPIHPKDFLSVM-----LEKSLS-AMRFLT-- 104  
 DB 50 PPXlpplr-----lkvpsnslq---ypagylgavpntneangsaameyltni 97  
 QY 105 -----AVNLEHP-EMLEKASRELWVRVMSRNEDITE-----PQSILAA 141

Db 98 lstkvydiaesplakklskrlgvmlykredlqpvsfklrgaymmvklpadqlak 157  
 QY 142 AERAGMSAEQAOQ-----LLEKIATPKVKNQKLTETEAACRYGAFGLPITVA 188  
 Db 158 gvicssagnhuqgvalsaskigtavivmpvttpeikwqavenl----- 201  
 QY 189 HVDGTHMLFGSDRMELLAHL---LGEKWMGPIPP 220  
 Db 202 ---gatvvlfgdsydaqahakiraeegitfipp 233

RESULT 6  
 Y32939  
 ID Y32939 standard; Protein; 592 AA.  
 XX Y32939;  
 AC  
 DT 09-NOV-1999 (first entry)  
 DE Mutant threonine dehydratase/deaminase protein sequence.  
 DE Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; mutin.  
 XX Arabidopsis thaliana.  
 OS Synthetic.  
 XX WO9941395-A1.  
 PN 19-AUG-1999.  
 PD 08-JAN-1999; 99WO-US00560.  
 PE 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 PA (PURD ) PURDUE RES FOUND.  
 XX Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;  
 XX WPI; 1999-527375/44.  
 DR N-PSDB; Z11197.  
 XX New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX Claim 13; Fig 8; 194pp; English.

This sequence represents a mutant Arabidopsis thaliana threonine dehydratase/deaminase (TD) protein of the invention. The protein is a feedback insensitive mutant. The TD DNA sequence is used as molecular marker (imparting resistance to toxic structural analogues of isoleucine) for selecting transformed cells and to produce transformants with increased levels of isoleucine (and thus better nutritional value) or of intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also TD-expressing plants permit use of the isoleucine structural analogues as herbicides. The DNA sequences are alternatives for antibiotic resistance markers (which are potentially harmful to the environment). Since no human analog of TD exists (humans can not synthesize isoleucine), it should be safe to use.

Sequence 592 AA;

Query Match 7.5%; Score 89; DB 20; Length 592;  
 Best Local Similarity 22.8%; Pred. No. 0.34;  
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;

us-08-978-174-1.rag

Thu Feb 15 15:59:12 2001

SQ Sequence 592 AA:  
 Query Match 7.5%; Score 89; DB 20; Length 592;  
 Best Local Similarity 22.8%; Pred. No. 0.34;  
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;  
 QY 55 PPGL-LPRKGLYMANDLKLRLHLLQIPHPKDFLSVM-----LEKGSLS-AMRFLT-- 104  
 Db 52 ppklplpr-----lkvspnslq-----ypagylgavpethneangsiacameyltni 99  
 QY 105 -----AVNLEHP-EMLEKASRELWVRVSRNEDITE-----PQSILAA 141  
 Db 100 lsktyvdiatpqlakklslrgvmylkredlpvfkfklrgaynmvklpadqlak 159  
 QY 142 AEKAGMSAEQAQG-----LLEKIATPKVNQLKETTEACRYGAFGLPIIVA 188  
 Db 160 gvicssagnhagvalsasklgtavimpvttpeikwqavenl----- 203  
 QY 189 HVDGQTHMLFGSDRMELLHL---LGEKWMGPPIP 220  
 Db 204 ---gatvvlfgdydaqahakiraeegltfipp 235  
 RESULT 8  
 Y32951  
 ID Y32951 standard; Protein; 592 AA.  
 XX AC Y32951;  
 XX AC 09-NOV-1999 (first entry)  
 DT DT Wild type threonine dehydratase/deaminase protein sequence.  
 DE DE Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 XX KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker.  
 XX OS Arabidopsis thaliana.  
 XX PN W09941395-A1.  
 XX PD 19-AUG-1999.  
 XX PF 08-JAN-1999; 99WO-US00560.  
 XX PR 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX PA (DOWC) DOW AGROSCIENCES LLC.  
 PA (PURD) PURDUE RES FOUND.  
 XX PI Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;  
 XX WPI; 1999-527375/44.  
 DR N-PSDB; Z11209.  
 XX PT New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX PS Disclosure; Page 86-89; 194pp; English.  
 CC This sequence is the wild type Arabidopsis thaliana threonine  
 CC dehydratase/deaminase (TD) protein. The invention relates to mutants of  
 CC the encoded protein, that are feedback insensitive TD mutants. The TD DNA  
 CC sequence is used as molecular marker (imparting resistance to toxic  
 CC structural analogues of isoleucine) for selecting transformed cells and  
 CC to produce transformants with increased levels of biosynthesis (and thus  
 CC better nutritional value) or of intermediates in biosynthesis of  
 CC isoleucine (e.g. 2-oxobutyrate, for synthesis of the degradable  
 CC biopolymer poly(hydroxybutyrate)). Also TD-expressing plants permit use

55 PPGL-LPRKGLYMANDLKLRLHLLQIPHPKDFLSVM-----LEKGSLS-AMRFLT-- 104  
 52 ppklplpr-----lkvspnslq-----ypagylgavpethneangsiacameyltni 99  
 105 -----AVNLEHP-EMLEKASRELWVRVSRNEDITE-----PQSILAA 141  
 100 lsktyvdiatpqlakklslrgvmylkredlpvfkfklrgaynmvklpadqlak 159  
 142 AEKAGMSAEQAQG-----LLEKIATPKVNQLKETTEACRYGAFGLPIIVA 188  
 160 gvicssagnhagvalsasklgtavimpvttpeikwqavenl----- 203  
 189 HVDGQTHMLFGSDRMELLHL---LGEKWMGPPIP 220  
 204 ---gatvvlfgdydaqahakiraeegltfipp 235  
 RESULT 7  
 Y32948  
 ID Y32948 standard; Protein; 592 AA.  
 XX AC Y32948;  
 XX AC 09-NOV-1999 (first entry)  
 DT DT Mutant threonine dehydratase/deaminase protein sequence.  
 DE DE Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 XX KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; mutcin.  
 XX OS Arabidopsis thaliana.  
 XX OS Synthetic.  
 XX PN W09941395-A1.  
 XX PD 19-AUG-1999.  
 XX PF 08-JAN-1999; 99WO-US00560.  
 XX PR 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX PA (DOWC) DOW AGROSCIENCES LLC.  
 PA (PURD) PURDUE RES FOUND.  
 XX PI Larrinua IM, Merlo DJ, Mourad GS, Pareddy DR;  
 XX WPI; 1999-527375/44.  
 DR N-PSDB; Z11206.  
 XX PT New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX PS Example 3; Page 126-129; 194pp; English.  
 CC This sequence represents a mutant Arabidopsis thaliana threonine  
 CC dehydratase/deaminase (TD) protein of the invention. The protein is a  
 CC feedback insensitive mutant. The TD DNA sequence is used as molecular  
 CC marker (imparting resistance to toxic structural analogues of isoleucine)  
 CC for selecting transformed cells and to produce transformants with  
 CC increased levels of isoleucine (and thus better nutritional value) or of  
 CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate; for  
 CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also  
 CC TD-expressing plants permit use of the isoleucine structural analogues as  
 CC herbicides. The DNA sequences are alternatives for antibiotic resistance  
 CC markers (which are potentially harmful to the environment). Since no  
 CC human analog of TD exists (humans can not synthesize isoleucine), it  
 CC should be safe to use.

of the isoleucine structural analogues as herbicides. The DNA sequences are alternatives for antibiotic resistance markers (which are potentially harmful to the environment). Since no human analog of ID exists (humans can not synthesize isoleucine), it should be safe to use.

```

Query Match          7.5%; Score 89; DB 20; Length 592;
Best Local Similarity 22.8%; Pred. NO. 0.34;
Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;

20Y 55 PPGL-LPRKGLYMANDKLLRHLHQLPIHPKPFQFLSYM-----LEKGSLS-AMRFLT-- 104
    || |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 52 Pkklplr-----lkvpsnlq-----ypaylgaavpertiaeangsaameynitnl 99

20Y 105 -----AVNLEHP-EMLEKASRELWNRVSRNEDITE-----POSTLAA 141
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 100 lstkvydaiesplglakklslrgvmylkredlqpvfsfklrgaynmvkvlpadqlak 159
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
20Y 142 AEKAGMSAEFAQG-----LLEKIAPKVKYNQLKTEACACRGAGLPTVA 188
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 160 gvicssagnhagvalsasklctavivmpttpelkwaqenl----- 203

189 HVDGOTHMFLGSDRMELLAAHL---LGEKWMGPPIP 220
    || |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 204 ---gatvvlfqdsydaqahakiraeeegltfipp 235
    || |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT	9
Y05702	
ID	Y05702 standard; Protein; 592 AA.
XX	
AC	Y05702;
XX	
DT	19-JUL-1999 (first entry)
XX	
DE	Arabidopsis wild-type threonine dehydratase/deaminase.
XX	
KW	Threonine dehydratase/deaminase; OMRI gene; feedback inhibition;
KW	transgenic plant; selectable marker; isoleucine.
XX	
OS	Arabidopsis thaliana.
XX	
FH	Key Location/Qualifiers
FT	Peptide 1..90
FT	/note= "transit peptide"
FT	Protein 91..592
FT	/note= "mature protein"
FT	Region 486..504
FT	/note= "regulatory region R4"
FT	Region 536..554
FT	/note= "regulatory region R6"
XX	
PN	WO9902656-A1.
XX	
PD	21-JAN-1999.
XX	
PF	10-JUL-1998; 98WO-US14362.
XX	
PR	17-FEB-1998; 98US-0074875.
PR	10-JUL-1997; 97US-0052096.
XX	
PA	(PURD ) PURDUE RES FOUND.
XX	
PI	Mourad GS;
XX	
DR	WPI; 1999-120860/10.
DR	N-PSDB; X25331.
XX	
PPT	New sequences encode mutant threonine dehydratase/deaminase - which
PPT	is insensitive to feedback inhibition, useful as a selective marker
PPT	to produce transformed cells resistant to toxic isoleucine analogues

Disclosure: Page 50-53; 120pp; English.

The present sequence represents *Arabidopsis thaliana* var. Columbia wild-type threonine dehydratase/deaminase (TD), the first enzyme of the isoleucine biosynthetic pathway. TD is encoded by the OMRI gene (see X25331) of *A. thaliana*. The invention provides nucleotide sequences (see X25332-40), originally isolated and cloned from *A. thaliana* mutated line Gm1lb (Omrl/omrl), which encode feedback insensitive TD that can be used to transform a wide variety of plants, fungi, bacteria and yeast. The mutated form of TD differs from the wild-type only by 2 point mutations (C to T at nucleotide 1955, and G to A at nucleotide 1631), which result in an R499C amino acid substitution in the regulatory region R4 of TD, and an R544H substitution in regulatory region R6. These forms of TD are not only insensitive to feedback inhibition by isoleucine, but are also insensitive to structural analogues of isoleucine that are toxic to plants and microorganisms which synthesize only wild-type TD. Nucleotide sequences encoding mutated forms of TD can therefore be used to create cells that are insensitive to compounds normally toxic to cells expressing only wild-type TD enzymes, and thus may be used to provide a biochemical selectable marker. Transformsants harboring a nucleotide sequence comprising a promoter operably linked to a mutated TD sequence demonstrate increased levels of isoleucine production, and thus provide an improved nutrient source.

Sequence 592 AA;

Query Match	7.5%	Score 89;	DB 20;	Length 592;
Best Local Similarity	22.8%;	Pred. No. 0.34;		
Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps				
QY	55	PPGL-LPRKGLYMANDLKLRLHLLQIPHFPEKDELFSVM-----LEKGSLS-AMRFLT--	104	
Dd			:    :	:
QY	52	ppklplr-----lkwpsnslq-----ypagyvgavpertneagngsaameyitni	95	
Dd			:    :	:
QY	105	-----AVNLFHP-EMLEKASRELWVMSRNEDITE-----PQSILAA	141	
Dd		: : : : : : : : : : :	: : : :	:
QY	100	lstkvydiatesplqtakiskrigvmymlkrediqpvfsklrgaynmvmklpadqlak	159	
Dd			: : : :	:
QY	142	AEKAGMSAEQAQG-----LLEKATPKVKLNOLKETTEACRGVAGGLPTIVA	188	
Dd		:     :	: :    :	:
QY	160	gvicssagnhaggvalssaklgctavimpvttpaikwqavenl-----	203	
Dd			:    :	:
QY	189	HVDGQTHMFGSDRMELLAHL---LGKWMGPFP	220	
Dd		: : : : :	: : : :	:
QY	204	----gatvvlfqdsydgaqhahakraeeegltfipp	235	
Dd			:    :	:

RESULT	10	
05703		
D	Y05703 standard; Protein; 592 AA.	
X		
X	Y05703;	
X		
X		
X		
X	19-JUL-1999 (first entry)	
X		
X	Feedback insensitive threonine dehydratase/de	
X		
X	Threonine dehydratase/deaminase; OMRI gene; f	
W	transgenic plant; selectable marker; isoleuc	
X		
X	Arabidopsis thaliana.	
X	Synthetic.	
X		
X		Location/Qualifiers
H	Key	1..90
H	Peptide	/note= "transit peptide"
T		91..592
T	Protéin	/note= "mature protein"
T		486..504
T	Region	/note= "regulatory region R4"
T		
T		

RESULT 11  
Y32952  
ID Y32952 standard; Protein; 600 AA.  
XX  
AC Y32952;  
XX  
DT 09-NOV-1999 (first entry)  
XX  
DE Mutant threonine dehydratase/deaminase protein sequence.  
XX  
KW Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
KW molecular marker; isoleucine toxic structural analog resistance;  
KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
KW polyhydroxybutyrate; antibiotic resistance marker; mutein.  
XX  
OS Arabidopsis thaliana.  
OS Synthetic.  
XX  
PN WO9941395-A1.  
XX  
PD 19-AUG-1999.  
XX  
PF 08-JAN-1999; 99WO-US00560.  
XX  
PR 10-JUL-1998; 98WO-US14362.  
PR 17-FEB-1998; 98US-0074875.  
XX  
PA (DOWC ) DOW AGROSCIENCES LLC.  
PA (PURD ) PURDUE RES FOUND.  
XX  
XX Larrinua IM, Marlo DJ, Mourad GS, Paredy DR;  
PI WPI; 1999-527375/44.  
XX N-PSDB; 211212.  
XX  
XX New nucleic acid encoding threonine dehydratase deaminase resistant  
PT to feedback inhibition, useful as selection marker for cell  
PT transformation and to impart herbicide resistance  
XX  
XX Example 2; Fig 5; 194pp; English.  
XX  
XX This sequence represents a mutant Arabidopsis thaliana threonine  
CC dehydratase/deaminase (TD) protein of the invention. The protein is a  
CC feedback insensitive mutant. The TD DNA sequence is used as molecular  
CC marker (imparting resistance to toxic structural analogues of isoleucine)  
CC for selecting transformed cells and to produce transformants with  
CC increased levels of isoleucine (and thus better nutritional value) or of  
CC intermediates in biosynthesis of isoleucine (e.g. 2-oxobutyrate, for  
CC synthesis of the degradable biopolymer poly(hydroxybutyrate)). Also  
CC TD-expressing plants permit use of the isoleucine structural analogues as  
CC herbicides. The DNA sequences are alternatives for antibiotic resistance  
CC markers (which are potentially harmful to the environment). Since no  
CC human analog of TD exists (humans can not synthesize isoleucine), it  
CC should be safe to use.  
XX  
XX Sequence 600 AA;  
SQ

Query Match 7.5%; Score 89; DB 20; Length 600;  
Best Local Similarity 22.8%; Pred. No. 0.35;  
Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;  
QY 55 PPGL-LPRKGLYMANDKLLRHLLQIPIHPKDFLSVM-----LEKGSLS-AMRFLT-- 104  
DB 52 ppklplpr-----lkvspnslq-----ypagylgavpntneangsaameyitni' 99  
QY 105 -----AVNLEHP-EMLEKASRELWMRVWSRNEDITE-----POSILAA 141  
DB 100 lskvydiaesplqlakklslrgvmylkredlqpvfskflrgaynmvklpadqlak 159  
QY 142 AEKAGMSAQAG-----LLEKIATPKVNOLKETTAAACRYGAGFLPIIVA 188  
DB 160 gvicssagahagqvalsasklgtavimvpttpaikwqavenl----- 203  
QY 189 HVDGOTMLFGSDRMELLAHL---LGEKWMGPPIP 220  
DB 204 ---gatvifgdsydaqahakiraeegitfipp 235

Region 536..554  
/note= "regulatory region R6"  
FT Misc-difference 499  
FT /note= "Arg in wild-type TD"  
FT Misc-difference 544  
FT /note= "Arg in wild-type TD"  
XX  
XX WO9902656-A1.  
XX  
XX 21-JAN-1999.  
XX  
XX 10-JUL-1998; 98WO-US14362.  
XX  
XX 17-FEB-1998; 98US-0074875.  
XX  
XX 10-JUL-1997; 97US-0052096.  
XX  
XX (PURD ) PURDUE RES FOUND.  
XX  
XX Mourad GS;  
XX  
XX WPI; 1999-120860/10.  
XX  
XX N-PSDB; X25332.  
XX  
XX New sequences encode mutant threonine dehydratase/deaminase - which  
PT is insensitive to feedback inhibition, useful as a selective marker  
PT to produce transformed cells resistant to toxic isoleucine analogues  
XX  
XX Disclosure; Page 53-56; 120pp; English.  
XX  
XX The present sequence represents an Arabidopsis thaliana mutant  
CC threonine dehydratase/deaminase (TD) precursor which, unlike  
CC wild-type TD, is insensitive to feedback inhibition by isoleucine.  
CC Claimed polynucleotides (see X2532-40), originally isolated and  
CC cloned from A. thaliana mutated line GM1b (omr1/omr1), encode  
CC feedback insensitive TD that can be used to transform a wide  
CC variety of plants, fungi, bacteria and yeast. Mutant TD differs  
CC from the wild-type enzyme only by an R499C amino acid substitution  
CC regulatory region R4, and by an R544H substitution in regulatory in  
CC region R6. Mutant TD is not only insensitive to feedback  
CC inhibition by isoleucine, but is also insensitive to structural  
CC analogues of isoleucine that are toxic to plants and microorganisms  
CC which synthesize only wild-type TD. Nucleotide sequences encoding  
CC mutated forms of TD can therefore be used to create cells that are  
CC insensitive to compounds normally toxic to cells expressing only  
CC wild-type TD enzymes, and thus may be used to provide a biochemical  
CC selectable marker. Transformants that express the mutant TD show  
CC increased levels of isoleucine production, and thus provide an  
CC improved nutrient source.  
XX  
XX Sequence 592 AA;  
SQ







CC structural analogues of isoleucine that are toxic to plants and  
 CC microorganisms which synthesize only wild-type TD. Nucleotide  
 CC sequences encoding mutated forms of TD can therefore be used to  
 CC create cells that are insensitive to compounds normally toxic to  
 CC cells expressing only wild-type TD enzymes, and thus may be used  
 CC to provide a biochemical selectable marker. Transformants  
 CC harboring a nucleotide sequence comprising a promoter operably  
 CC linked to a mutated TD sequence demonstrate increased levels of  
 CC isoleucine production, and thus provide an improved nutrient source.  
 XX  
 SQ Sequence 609 AA;

Query Match 7.5%; Score 89; DB 20; Length 609;  
 Best Local Similarity 22.8%; Pred. NO. 0.36; 57; Indels 80; Gaps 11;  
 Matches 49; Conservative 29; Mismatches 57; Indels 80; Gaps 11;  
 QY 55 PPGL-LPRKGLYMANDKLLRHHLQIPHPKDFLSVM-----LEKGLS-AMRFLT-- 104  
 DB 69 pklplpr-----lkvspnslq---ypagylgavperrtneangsiaameyltni 116  
 QY 105 -----AVNLEHP-EMLEKASRELMMRVMSRNEDITE-----POSILAA 141  
 DB 117 lskvydiaesplqlaklsklrgvmyikredlqpvsfklrgaynmwklpadqlak 176  
 QY 142 AEKAGMSAQAG-----LLEKIATPKVNQKLTETEAACRYGAFGLPITVA 188  
 DB 177 gvicssagnhagdvaisaaklgctavimpvttpeikwqvenl----- 220  
 QY 189 HVDGOTMLFGSDRMELLALH---LGEKWMGPIPP 220  
 DB 221 ---gatvlfsgsydaqahakiraeegltfipp 252

RESULT 14  
 W04266  
 ID W04266 standard; Protein; 308 AA.  
 AC W04266;  
 XX 30-JUN-1997 (first entry)  
 XX Inosine-guanosine kinase.  
 XX Inosine-guanosine kinase; 5'-inosinic acid; 5'-guanylic acid;  
 KW Corynebacterium ammoniagenes; ATP; seasoning; food.  
 XX Exiguobacterium sp. (ATCC 35652).  
 OS W09630501-A1.  
 PN 03-OCT-1996.  
 XX 22-MAR-1996; 96WO-JP00761.  
 XX 09-JUN-1995; 95JP-0177900.  
 PR 24-MAR-1995; 95JP-0102888.  
 XX (AJIN ) AJINOMOTO CO INC.  
 PA Kawasaki H, Shimaoka M, Usuda Y, Utagawa T;  
 PI WPI: 1996-455349/45.  
 DR N-PSDB; T33972.  
 XX Prodn. of 5'-inosinic acid or 5'-guanylic acid - using C.  
 PT ammoniagenes transformed by inosine-guanosine kinase gene, for use,  
 PT in food seasoning  
 XX Example 14; Page 65-66; 72pp; Japanese.  
 PS This sequence represents the inosine-guanosine kinase sequence derived  
 CC from Exiguobacterium sp. This sequence was used in the production of

CC 5'-inosinic acid or 5'-guanylic acid by reacting an organism, pref.  
 CC Corynebacterium ammoniagenes, capable of regenerating ATP, containing  
 CC this recombinant DNA, with inosine, guanosine or their precursors, and  
 CC an energy and phosphoric acid source. 5'-inosinic acid and 5'-guanylic  
 CC acid are useful as seasonings for food.  
 XX  
 SQ Sequence 308 AA;

Query Match 7.0%; Score 83.5; DB 17; Length 308;  
 Best Local Similarity 25.8%; Pred. NO. 0.52; 57; Indels 33; Gaps 8;  
 Matches 40; Conservative 25; Mismatches 25; Indels 33; Gaps 8;  
 QY 63 GLYVANDKLLRHHLQIPHPKDFLSVMLEKGLS-AMRFLTAVNLEHPEMLEKASRELW 122  
 DB 40 grnvaqlgvlgnvdrfvstvtndqgi-----gvleelrsln-vnvehvdilledngmgw 94  
 QY 123 MRVMSRNEDITEPQSILAAAEKAGMSAQAGLLEKIATPKVK-----NOLKE 170  
 DB 95 lavmdnngdl--qtsiskqpdeame---qcilrridvfaestavaldidlsnvlne 148  
 QY 171 TTEACR-----YGAFGGLPITVAHVVDGQTHMLFG 199  
 DB 149 tie-lcremkpilygvvgc---hlsviernrhliqg 179

RESULT 15  
 Y32942  
 ID Y32942 standard; Protein; 539 AA.  
 AC Y32942;  
 XX 09-NOV-1999 (first entry)  
 XX Mutant threonine dehydratase/deaminase protein sequence.  
 DE Threonine dehydratase/deaminase; TD; feedback insensitive mutant;  
 KW molecular marker; isoleucine toxic structural analog resistance;  
 KW isoleucine production; biosynthesis; degradable biopolymer; herbicide;  
 KW polyhydroxybutyrate; antibiotic resistance marker; muten.  
 XX Arabidopsis thaliana.  
 OS Synthetic.  
 XX W09941395-A1.  
 PN 19-AUG-1999.  
 PD 08-JAN-1999; 99WO-US00560.  
 XX 10-JUL-1998; 98WO-US14362.  
 PR 17-FEB-1998; 98US-0074875.  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 PA (PURD ) PURDUE RES FOUND.  
 XX Larrinua IM, Merlo DU, Mourad GS, Paredy DR;  
 DR WPI: 1999-527375/44.  
 DR N-PSDB; 211200.  
 XX New nucleic acid encoding threonine dehydratase deaminase resistant  
 PT to feedback inhibition, useful as selection marker for cell  
 PT transformation and to impart herbicide resistance  
 XX Claim 13; Page 106-109; 194pp; English.  
 PS This sequence represents a mutant Arabidopsis thaliana threonine  
 CC dehydratase/deaminase (TD) protein of the invention. The protein is a  
 CC feedback insensitive mutant. The TD DNA sequence is used as molecular  
 CC marker (imparting resistance to toxic structural analogues of isoleucine)  
 CC for selecting transformed cells and to produce transformants with  
 CC increased levels of isoleucine (and thus better nutritional value) or of

Sequence 539 AA;

Search completed: February 15, 2001, 15:52:05  
Job time: 83 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2001, 15:50:48 ; Search time 14.15 Seconds  
(without alignments)  
286.805 Million cell updates/sec

Title: US-08-978-174-1  
Perfect score: 1185  
Sequence: 1 MGPLPRTVELFDVLSYV.....AHLGKWMGPPIPAVNRL 226

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues  
Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1185	100.0	226	3	US-08-978-174-1
2	874	73.8	226	3	US-08-978-174-3
3	80.5	6.8	339	2	US-08-855-714-3
4	76.5	6.5	317	2	US-09-066-075-2
5	76.5	6.5	317	2	US-08-518-615A-2
6	76.5	6.5	317	3	US-08-951-889-2
7	76.5	6.5	497	1	US-08-075-193-4
8	76.5	6.5	497	2	US-08-564-090A-4
9	76.5	6.5	497	4	PCT-US94-06698-4
10	75.5	6.4	1398	1	US-08-750-532-9
11	73.5	6.2	523	2	US-08-473-553A-3
12	73.5	6.2	869	1	US-08-188-582-32
13	73.5	6.2	869	1	US-08-646-715-32
14	73.5	6.2	980	2	US-08-473-553A-6
15	73.5	6.2	985	2	US-08-473-553A-2
16	73	6.2	948	1	US-08-698-551-14
17	73	6.2	948	2	US-08-602-228-14
18	73	6.2	948	2	US-08-533-901B-14
19	73	6.2	948	2	US-08-839-032A-14
20	73	6.2	948	2	US-08-839-031A-14
21	73	6.2	948	4	PCT-US95-12724-14
22	71.5	6.0	543	2	US-08-922-170B-10
23	71	6.0	587	1	US-07-955-905A-23
24	70.5	5.9	615	2	US-08-484-101B-38
25	70	5.9	401	1	US-08-198-446B-11
26	70	5.9	401	2	US-08-870-693-11
27	69.5	5.9	259	2	US-07-857-224B-51
28	69.5	5.9	529	4	PCT-US95-0500B-15

Sequence 2, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 10, Appli  
Sequence 50, Appli  
Sequence 14, Appli  
Sequence 4, Appli  
Sequence 4, Appli

## ALIGNMENTS

RESULT 1  
US-08-978-174-1  
; Sequence 1, Application US/08978174  
; Patent No. 6030809  
; GENERAL INFORMATION:  
; APPLICANT: Shah, Purvi  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: NEW GLUTATHIONE-S-TRANSFERASE  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/978,174  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0430 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 226 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: BLADTUT04  
; CLONE: 1554593  
US-08-978-174-1

Query Match 100.0% Score 1185; DB 3; Length 226;

Best Local Similarity 100.0%; Pred. No. 4.3e-126;  
Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGPLPRTVELFDVLSYSLGFEILCRQYQNIWNLQRLPSLITGIMKDSGNKPPGLLP 60  
Db 1 MGPLPRTVELFDVLSYSLGFEILCRQYQNIWNLQRLPSLITGIMKDSGNKPPGLLP 60  
QY 61 RGLYMANDLKLRLHHLQIPHFDPKFLSVMLKGSLSAMRFLTAVNLEHPMLEKASRE 120  
Db 61 RGLYMANDLKLRLHHLQIPHFDPKFLSVMLKGSLSAMRFLTAVNLEHPMLEKASRE 120  
QY 121 LMRVWSRNEDETEPOSILAAAEKAGMSAQOGLLEKIATPKYKNQKJETTEAACRYGA 180  
Db 121 LMRVWSRNEDETEPOSILAAAEKAGMSAQOGLLEKIATPKYKNQKJETTEAACRYGA 180  
QY 181 FGLPTTVAHVGDGTHMLFGSDRMELLALLGKWMGPIPPAVNARL 226  
Db 181 FGLPTTVAHVGDGTHMLFGSDRMELLALLGKWMGPIPPAVNARL 226

RESULT 2  
US-08-978-174-3  
Sequence 3, Application US/08978174  
Patent No. 6030809  
GENERAL INFORMATION:  
APPLICANT: Shah, Purvi  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: NEW GLUTATHIONE-S-TRANSFERASE  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/978,174  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0430 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-845-0555  
TELEFAX: 650-845-4166  
TELEX:

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 226 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: ?  
US-08-978-174-3

Query Match 73.8%; Score 874; DB 3; Length 226;  
Best Local Similarity 69.5%; Pred. No. 6.2e-91;  
Matches 157; Conservative 38; Mismatches 31; Indels 0; Gaps 0;

QY 1 MGPLPRTVELFDVLSYSLGFEILCRQYQNIWNLQRLPSLITGIMKDSGNKPPGLLP 60  
Db 1 MGPLPRTVELFDVLSYSLGFEILCRQYQNIWNLQRLPSLITGIMKDSGNKPPAMYP 60  
QY 61 RGLYMANDLKLRLHHLQIPHFDPKFLSVMLKGSLSAMRFLTAVNLEHPMLEKASRE 120  
Db 61 HKQYILKEIPLKQLFQVPMSPKDFGHEHVKGTVNMRFLTAVSMEQPEMLEKVSRE 120  
QY 121 LMRVWSRNEDETEPOSILAAAEKAGMSAQOGLLEKIATPKYKNQKJETTEAACRYGA 180  
Db 121 LMRVWSRNEDETEPOSILAAAEKAGMSAQOGLLEKIATPKYKNQKJETTEAACRYGA 180  
QY 181 FGLPTTVAHVGDGTHMLFGSDRMELLALLGKWMGPIPPAVNARL 226  
Db 181 FGLPTTVAHVGDGTHMLFGSDRMELLALLGKWMGPIPPAVNARL 226

RESULT 3  
US-08-855-714-3  
Sequence 3, Application US/08855714  
Patent No. 5939075  
GENERAL INFORMATION:  
APPLICANT: Houng, Huo-Shu H.  
APPLICANT: Warren, Richard L.  
TITLE OF INVENTION: MUTANTS OF BRUCELLA MELITENSIS  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John Moran, Esq.  
STREET: HQ USAMRDC, Dept. of Army, Fort Detrick  
CITY: Frederick  
STATE: MD  
COUNTRY: US  
ZIP: 21702-5012  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/855,714  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/334,129  
FILING DATE: 04-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Hendricks, Glenn  
REGISTRATION NUMBER: 32,535  
REFERENCE/DOCKET NUMBER: 08/143,692  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 619-2065  
TELEFAX: (301) 619-7714  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 339 amino acids  
TYPE: amino acid  
STRANDEDNESS: both  
TOPOLOGY: unknown  
MOLECULE TYPE: NO  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: BRUCELLA MELITENSIS  
US-08-855-714-3

Query Match 6.8%; Score 80.5; DB 2; Length 339;  
Best Local Similarity 21.7%; Pred. No. 0.59;  
Matches 65; Conservative 25; Mismatches 96; Indels 113; Gaps 14;  
QY 3 PLPRTVELFDVLSYSLGFEILCRQYQNIWNLQRLPSLITGIMKDSGNKPPGLLP 59

Db 47 PLPTSPMLKMCQSAPPTSWLKRRLFCPPRPHWK-----SLRTASQKSSFTKAALK 99  
Qy 60 PRK-GLYMANDLKLRLHQLQIPHFQKDFLSVMLEKGS-----SAMRELTAVNLEH 110  
Db 100 PRPGSWTRKRSSPSALGRANH-PQDRRLGYDGKGQVRLASLDETQACNAFAAIN-KA 157  
Qy 111 PEMLE---KASRELHMRVMSRNEDITEPQSIILAAAKAG-----TPKVNQLKETTE 173  
Db 158 PAILEGFVEFEV-----SVIAARDRSGNVAIFDLAENVHKGILATST 202  
Qy 147 -----MSAEQAQGLLEKTA-----TPKVNQLKETTE 173  
Db 203 VPAAISVQTAEAARTAAEKHLHALDYGVLGFEFVLKDGTLTLANEFAPRVHS-GHWTE 261  
Qy 174 AACRYGAF-----GLPT--TVAHVDGOTHMLFGSD-----RMELLAHLGEX 213  
Db 262 ACAISQFEQHRAVAGLPLGNTDRHSDCVNENLIGDIEKVPAILCEKNVHLHYGKK 320

RESULT 4  
US-09-066-075-2  
Sequence 2, Application US/09066075  
Patent No. 5925749  
GENERAL INFORMATION:  
APPLICANT: Mathur, E., et al.  
TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,075  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/518,615  
FILING DATE: August 23, 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 331400-20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 317 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS:  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-09-066-075-2

Query Match 6.5%; Score 76.5; DB 2; Length 317;  
Best Local Similarity 21.1%; Pred. No. 1.5;  
Matches 43; Conservative 22; Mismatches 50; Indels 89; Gaps 10;  
Qy 76 HLQIPH-----FP-----KDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEK 116  
Db 49 HVRIPRWSTHAYAFPPYKIMDRFFKRVDEINGALKRG-----LAVAINIHHELMN 102  
Qy 117 ASRE-----LWMRVMSRNE-----ITEPQS-----137

Db 103 DPEEHKERFLAWKQIADRYKDYPTETLFFELLNPHGNLTPEKWNELLEALKVIRSIDK 162  
Qy 138 ---TLAAAEKAGMSAEQAQGLLEKTIATPKVNQLKETTEACACRYGAGLPTVAHVQDG 193  
Db 163 KHTIIIGTAEWGGISA-----LEKUSVPKWE---KNSIVTIHYNPFEF-----203  
Qy 194 THMLFGSDRMELLAHLGKWMGP 217  
Db 204 THQ--GAWEVGESEKWLGRKWGSP 225

RESULT 5  
US-08-518-615A-2  
Sequence 2, Application US/08518615A  
Patent No. 5962258  
GENERAL INFORMATION:  
APPLICANT: Mathur, E., et al.  
TITLE OF INVENTION: Carboxymethyl Cellulase from Thermotoga Maritima  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/518,615A  
FILING DATE: August 23, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 331400-20  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 317 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS:  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-08-518-615A-2

Query Match 6.5%; Score 76.5; DB 2; Length 317;  
Best Local Similarity 21.1%; Pred. No. 1.5;  
Matches 43; Conservative 22; Mismatches 50; Indels 89; Gaps 10;  
Qy 76 HLQIPH-----FP-----KDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEK 116  
Db 49 HVRIPRWSTHAYAFPPYKIMDRFFKRVDEINGALKRG-----LAVAINIHHELMN 104  
Qy 117 ASRE-----LWMRVMSRNE-----ITEPQS-----137  
Db 103 DPEEHKERFLAWKQIADRYKDYPTETLFFELLNPHGNLTPEKWNELLEALKVIRSIDK 162  
Qy 138 ---TLAAAEKAGMSAEQAQGLLEKTIATPKVNQLKETTEACACRYGAGLPTVAHVQDG 193  
Db 163 KHTIIIGTAEWGGISA-----LEKUSVPKWE---KNSIVTIHYNPFEF-----203

QY 194 THMLFGSDRMELLALLGKWMGP 217  
|||:::|  
Db 204 THQ--GAEWVEGSEKWLGRKMGSP 225

RESULT 6  
US-08-951-889-2  
; Sequence 2, Application US/08951889  
; Patent No. 6008032  
; GENERAL INFORMATION:  
; APPLICANT: Mathur, E., et al.  
; TITLE OF INVENTION: Carboxymethyl Cellulase from  
; TITLE OF INVENTION: Thermotoga Maritima  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN,  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/951,889  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/518,615  
; FILING DATE: August 23, 1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FERRARO, GREGORY D.  
; REGISTRATION NUMBER: 36,134  
; REFERENCE/DOCKET NUMBER: 331400-20  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 317 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-951-889-2

Query Match 6.5%; Score 76.5; DB 3; Length 317;  
Best Local Similarity 21.1%; Pred. No. 1.5;  
Matches 43; Conservative 22; Mismatches 50; Indels 89; Gaps 10;

QY 76 HLQIPH-----PP-----KDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEK 116  
|||:::|  
Db 49 HVRIPIWSTHAYAPPYKIMDRFFKRVDEVINGALKRG-----LAVAINIHVEELMN 102  
QY 117 ASRE-----LWVRVSRNED-----ITEPQS-----137  
|||:::|  
Db 103 DPEHKEERFLALWQIADRYKDYPTFLFEILNEPHGNLPKKNWELLEBAKLVIRSTDK 162  
QY 138 -----ILAAAEKAGMSAEQAQGLLEKATPKVKKNOLKETTETAAACRYGAFGLPITVAHVDGQ 193  
|||:::|  
Db 163 KHTIIIGTAEWGGISA-----LEKLSVPKWE---KNSIVTIHYNPFEE-----203  
QY 194 THMLFGSDRMELLALLGKWMGP 217  
|||:::|  
Db 204 THQ--GAEWVEGSEKWLGRKMGSP 225  
RESULT 7

US-08-075-193-4  
; Sequence 4, Application US/08075193  
; Patent No. 5547868  
; GENERAL INFORMATION:  
; APPLICANT: MILLER, WALTER L.  
; APPLICANT: HARIKRISHNA, JENNIFER A.  
; TITLE OF INVENTION: CHOLESTEROL DISPOSAL FUSION ENZYMES  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM  
; STREET: FIVE PALO ALTO SQUARE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/075,193  
; FILING DATE: 09-JUN-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: NEELEY PH.D., RICHARD L.  
; REGISTRATION NUMBER: 30,092  
; REFERENCE/DOCKET NUMBER: UCAL-236/00US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-494-7622  
; TELEFAX: 415-857-0663  
; TELEX: 380816 COOLEY PA  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 497 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-075-193-4

Query Match 6.5%; Score 76.5; DB 1; Length 497;  
Best Local Similarity 25.4%; Pred. No. 3;  
Matches 44; Conservative 29; Mismatches 69; Indels 31; Gaps 9;

QY 14 VLSPYSWLGFELCRYONWNNQLRPSLITGIMKDSGNKPPGLPRKG-LYMANDKL 72  
|||:::|  
Db 195 LITPPEHLEALLCQRTDITKAAL-----GVLRQSRVKVWLVGRRGPLQVAFPIKE 246  
QY 73 LRHHLQI-----PIHFPKDFL-----SVMLEKGSLSAMRFLTAVNLEHPEMLEKASREL 121  
|||:::|  
Db 247 LREMILQPGARPLDLPVDFLGLQDKIKEVPRPRKRLTELLRTAT--EKPCPAEAARQAS 304  
QY 122 WNRVSRNEDLITEPQSILAAAEKAGMSAEQAQGLLEKATPKVKKNOLKETTEA 174  
|||:::|  
Db 305 ASRANGL-RFRSPQOVLPSPD-----GRRAGV--RLAVTRLEG-VDEATRA 348

RESULT 8  
US-08-564-090A-4  
; Sequence 4, Application US/08564090A  
; Patent No. 5939318  
; Patent No. 5939318 5741703  
; GENERAL INFORMATION:  
; APPLICANT: MILLER, WALTER L.  
; APPLICANT: HARIKRISHNA, JENNIFER A.  
; TITLE OF INVENTION: CHOLESTEROL DISPOSAL FUSION ENZYMES  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOLEY GODWARD LLP  
; STREET: FIVE PALO ALTO SQUARE

STREET: 3000 EL CAMINO REAL  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/364,090A  
FILING DATE: 02/05/96  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD L. NEELEY, PH.D.  
REGISTRATION NUMBER: 30,092  
REFERENCE/DOCKET NUMBER: UCAL-236/01US  
TELEPHONE: 415-843-5000  
TELEFAX: 415-857-0663  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 497 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-564-090A-4

Query Match 6.5%; Score 76.5; DB 2; Length 497;  
Best Local Similarity 25.4%; Pred. No. 3;  
Matches 44; Conservative 29; Mismatches 69; Indels 31; Gaps 9;  
QY 14 VLSPSYWLGFELCRQYQNIWNLQRLPSLITGIMKDSGNKPPGLPRKG-LYMANDLKL 72  
DB 195 LLTPPEHLEALLCQRTDITKAAL-----GVLRQSRVKTVMVLRGGRPLQVAFITKE 246  
QY 73 LRHLQI-----PIHFKDFL-----SYMLEKGSLSAMRFLTAVNLEHPEMLEKASREL 121  
DB 247 LREMIQPGARPILDVDFLGLQDKIKEVPRPKRTELLLRAT--EKPGPAEAAQAS 304  
QY 122 WMRVSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKYKNQKTEEA 174  
DB 305 ASRAWGL-REFRSPQVLPSPD-----GRRAGV--RLAVTRLEG-VDEATRA 348

RESULT 9  
PCT-US94-06698-4  
Sequence 4, Application PC/TUS9406698  
GENERAL INFORMATION:  
APPLICANT: MILLER, WALTER L.  
APPLICANT: HARIKRISHNA, JENNIFER A.  
TITLE OF INVENTION: CHOLESTEROL DISPOSAL FUSION ENZYMES  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ROBBINS, BERLINER & CARSON  
STREET: 201 NORTH FIGUEROA STREET  
CITY: LOS ANGELES  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 90012  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/06698  
FILING DATE: FILED HERewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

NAME: BERLINER, ROBERT  
REGISTRATION NUMBER: 20,121  
REFERENCE/DOCKET NUMBER: 5555-224-C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 213-977-1001  
TELEFAX: 213-977-1003  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 497 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-06698-4

Query Match 6.5%; Score 76.5; DB 4; Length 497;  
Best Local Similarity 25.4%; Pred. No. 3;  
Matches 44; Conservative 29; Mismatches 69; Indels 31; Gaps 9;  
QY 14 VLSPSYWLGFELCRQYQNIWNLQRLPSLITGIMKDSGNKPPGLPRKG-LYMANDLKL 72  
DB 195 LLTPPEHLEALLCQRTDITKAAL-----GVLRQSRVKTVMVLRGGRPLQVAFITKE 246  
QY 73 LRHLQI-----PIHFKDFL-----SYMLEKGSLSAMRFLTAVNLEHPEMLEKASREL 121  
DB 247 LREMIQPGARPILDVDFLGLQDKIKEVPRPKRTELLLRAT--EKPGPAEAAQAS 304  
QY 122 WMRVSRNEDITEPQSILAAAEKAGMSAEQAQGLLEKIATPKYKNQKTEEA 174  
DB 305 ASRAWGL-REFRSPQVLPSPD-----GRRAGV--RLAVTRLEG-VDEATRA 348

RESULT 10  
US-08-750-532-9  
Sequence 9, Application US/08750532  
Patent No. 5756339  
GENERAL INFORMATION:  
APPLICANT: MITTA, Masanori  
APPLICANT: YAMAMOTO, Katsuhiko  
APPLICANT: MORISHITA, Mio  
APPLICANT: ASADA, Kiyozo  
APPLICANT: TSUNASAWA, Susumu  
APPLICANT: KATO, Ikunoshin  
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.  
STREET: 419 Seventh Street N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: United States of America  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/750,532  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JJP95/01095  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 1994/130236  
FILING DATE: 13-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 1994/173912  
FILING DATE: 26-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: MITTA-1  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1398 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-750-532-9

Query Match 6.4%; Score 75.5; DB 1; Length 1398;  
Best Local Similarity 16.8%; Pred. No. 19;  
Matches 39; Conservative 39; Mismatches 69; Indels 85; Gaps 9;  
QY 50 DSGNKPGLPRKGLY--MANDLKLRLHHLQI----- 79  
DB 1097 DKADFAGVLTPAEGVLGEARNYTLIVKHALTLEVPVPNATVIIGNTYLTLDENGTVTFYA 1156  
QY 80 PIHPKDFLSVLMKGLSLSAMRELTAVNLEHPMELEKASRELWVRWNRNEDITEPQSIL 139  
DB 1157 PTKLGSDEITVIVKKNFNTLEKTFQITVSEPEITE-----EDINEPKLAM 1202  
QY 140 AAEEKAG--MSAE-QAOGLEKIAT-----PKVKNQLKETTAAACR 177  
DB 1203 SSPEANATIVSEMESEGGVKTVITVEITTINGTANETATIVVPVKKAEINIEVSGDHVIS 1262  
QY 178 YG-----AFGLPIITVAHVGDQTHMLFGSDRMELLA-HLGEKW 214  
DB 1263 YSIEBEGYAKYVITVKFASPTVT-----VVTYTIYAGPRVSILTLENFLGYSW 1310

RESULT 11  
US-08-473-553A-3  
Sequence 3, Application US/08473553A  
Patent No. 5859338  
GENERAL INFORMATION:  
APPLICANT: Meyerowitz, Elliot M.  
APPLICANT: Clark, Steven E.  
APPLICANT: Williams, Robert W.  
TITLE OF INVENTION: Plant Clavatal Nucleic Acids,  
TITLE OF INVENTION: Transformed Plants, and Proteins  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/473,553A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-60886/RFT/RMS  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 523 amino acids

TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-473-553A-3

Query Match 6.2%; Score 73.5; DB 2; Length 523;  
Best Local Similarity 21.8%; Pred. No. 7;  
Matches 52; Conservative 34; Mismatches 77; Indels 75; Gaps 14;  
QY 36 NQLRPSLTGIMKDSGNKPP---GLPRKGLYMANDLKLRLHHLQIPIHPKDFLSVNL 92  
DB 194 NLKHLHTFLHNNLTGHIPPGLSVLSKSL-----DLST-----NQLTGEIPQSFNL-- 243  
QY 93 EKGSLSAMRFTAVNL-----EHPMELEKASRELWVRWNRNEDITEPOSILAAAE-- 143  
DB 244 --GN-----ITLINFRNNLYGQIPEAIGELPKLEFVFWENNFTLQLPANLGRGNLI 295  
QY 144 KAGMSAEQAQGLL-----EKIATPK-----VKNLKET 171  
DB 296 KLDVSDNHLTKLPKDLGRGEKLEMLILSNNEFFGPIPELGEKCKSLTKIRIVKNLLNGT 355  
QY 172 TEACRYGARGCLP-ITVAHV-----DCQTHMLFGSDRMELLALLLGEKWM-GPIPPAV 222  
DB 356 VPA-----GLFNLPLVTIETLDTNFFSGELPVTMSGVDVLDQI---YLSNNFSGEIPPAI 407

RESULT 12  
US-08-188-582-32  
Sequence 32, Application US/08188582  
Patent No. 5534410  
GENERAL INFORMATION:  
APPLICANT: Tjian, Robert  
APPLICANT: Comai, Lucio  
APPLICANT: Dynlacht, Brian D.  
APPLICANT: Hoey, Timothy  
APPLICANT: Ruppert, Siegfried  
APPLICANT: Tanese, Naoko  
APPLICANT: Wang, Edith  
APPLICANT: Weinzierl, Robert O.J.  
TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/188,582  
FILING DATE: 28-JAN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 869 amino acids  
TYPE: amino acid  
TOPOLOGY: linear



MOLECULE TYPE: protein  
US-08-188-582-32

Query Match 6.2%; Score 73.5; DB 1; Length 869;  
Best Local Similarity 20.5%; Pred. No. 15;  
Matches 44; Conservative 30; Mismatches 82; Indels 59; Gaps 10;

QY 32 IWNINLQRPSTLITIMKDSGNKPPCLLPKGLYMANDKLLRHHLQIPIHFPPKDFLSVM 91  
DB 318 VMKFGKQWQPTLLQAMQVEKGATGSLP-----HLPGLAICSRSGAVCLMSPEDGLRQI 373  
QY 92 LEKGSLSAMRFLTAVNLEHPMLEKASRELWMRVWSRNEDETEPQSILAAAEKAGMSAEQ 151  
DB 374 -----YRDPETLVFRDSSWR--WA---DFTAHPRLVTGDRGTGVMKMD 412  
QY 152 AQG-----LLEKIATPKVKNLKETEACRYGA-----FG-----LPITVAHVDG 192  
DB 413 TQGPFGCGLLFRIG-----AEASCQKGERVLLTQYLGHSPPKCLPPTLHLVCT 461  
QY 193 QTHMLFGSDRMELLALLGEKWMGPPI-PAVNARL 226  
DB 462 QFSLYLVDRLPLVPML-----KWNHGLPSPLLARL 493

RESULT 13

US-08-646-715-32  
Sequence 32, Application US/08646715  
Patent No. 5637886

GENERAL INFORMATION:  
APPLICANT: Tjian, Robert  
APPLICANT: Comai, Lucio  
APPLICANT: Dynlacht, Brian D.  
APPLICANT: Hoey, Timothy  
APPLICANT: Ruppert, Siegfried  
APPLICANT: Tanese, Naoko  
APPLICANT: Wang, Edith  
APPLICANT: Weinzierl, Robert O. J.

TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE  
NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESS: FLEHR, HOBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/646,715  
FILING DATE: 09-MAY-1996  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/188,582  
FILING DATE: 28-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard A.  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-57650-2/ATT/RAO

TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 869 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-646-715-32

Query Match 6.2%; Score 73.5; DB 1; Length 869;  
Best Local Similarity 20.5%; Pred. No. 15;  
Matches 44; Conservative 30; Mismatches 82; Indels 59; Gaps 10;

QY 32 IWNINLQRPSTLITIMKDSGNKPPCLLPKGLYMANDKLLRHHLQIPIHFPPKDFLSVM 91  
DB 318 VMKFGKQWQPTLLQAMQVEKGATGSLP-----HLPGLAICSRSGAVCLMSPEDGLRQI 373  
QY 92 LEKGSLSAMRFLTAVNLEHPMLEKASRELWMRVWSRNEDETEPQSILAAAEKAGMSAEQ 151  
DB 374 -----YRDPETLVFRDSSWR--WA---DFTAHPRLVTGDRGTGVMKMD 412  
QY 152 AQG-----LLEKIATPKVKNLKETEACRYGA-----FG-----LPITVAHVDG 192  
DB 413 TQGPFGCGLLFRIG-----AEASCQKGERVLLTQYLGHSPPKCLPPTLHLVCT 461  
QY 193 QTHMLFGSDRMELLALLGEKWMGPPI-PAVNARL 226  
DB 462 QFSLYLVDRLPLVPML-----KWNHGLPSPLLARL 493

RESULT 14

US-08-473-553A-6  
Sequence 6, Application US/08473553A  
Patent No. 5859338

GENERAL INFORMATION:  
APPLICANT: Meyerowitz, Elliot M.  
APPLICANT: Clark, Steven E.  
APPLICANT: Williams, Robert W.

TITLE OF INVENTION: Plant Clavatal Nucleic Acids,  
TITLE OF INVENTION: Transformed Plants, and Proteins  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESS: Flehr, Hobbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/473,553A  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-60886/RFT/RMS

TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 980 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-473-553A-6

Query Match 6.2%; Score 73.5; DB 2; Length 980;  
Best Local Similarity 21.8%; Pred. No. 18;  
Matches 52; Conservative 34; Mismatches 77; Indels 75; Gaps 10;



Thu Feb 15 15:59:13 2001

us-08-978-174-1.rai

---

Page 9



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2001, 15:50:48 ; Search time 15:51 Seconds  
(without alignments)  
989.397 Million cell updates/sec

Title: US-08-978-174-1

Perfect score: 1185  
Sequence: 1 MGPLPRTVELFYDVLSPYSW.....AHLLEKMGKPIPPAVNARL 226

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: pir1.\*
- 2: pir2.\*
- 3: pir3.\*
- 4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	328.5	27.7	226	T27747	hypothetical prote
2	298.5	25.2	225	T34201	hypothetical prote
3	198	16.7	195	G83629	hypothetical prote
4	164.5	13.9	194	S72164	2-hydroxychromene-
5	140	11.8	197	T31286	2-nitrotoluene dio
6	125	10.5	33	S17164	glutathione trans
7	95	8.0	203	C55552	2-hydroxychromene-
8	91	7.7	581	H72425	ABC transporter A
9	90.5	7.6	410	G69127	flavoprotein A hom
10	89.5	7.6	962	1 SNECFI	patrilysin (EC 3.4
11	89	7.5	199	T49343	probable isomerase
12	89	7.5	592	T51712	threonine dehydrat
13	88.5	7.5	613	A35296	secretogranin II p
14	87	7.3	926	B83375	probable glycosyl
15	84.5	7.1	316	B71301	probable tRNA delt
16	84.5	7.1	971	2 S54595	probable membrane
17	83.5	7.0	488	2 A40367	catalase (EC 1.1.1.
18	83.5	7.0	700	1 S09748	kinesin-related pr
19	82.5	7.0	263	T51169	hypothetical prote
20	81.5	6.9	619	2 S02180	secretogranin II -
21	81.5	6.9	1061	1 D7AB12	DNA-directed DNA p
22	81	6.8	274	2 J55896	hypothetical prote
23	80.5	6.8	705	2 A48144	protein kinase CBC
24	80	6.8	402	2 E69107	hypothetical prote
25	80	6.8	540	2 A70358	topoisomerase I -
26	79.5	6.7	547	2 E69647	catalase (EC 1.1.1.
27	79.5	6.7	1073	1 QYHUX	heat-stable entero
28	79	6.7	230	2 C75301	rRNA methylase Spo
29	78.5	6.6	333	2 F83215	conserved hypothet

## ALIGNMENTS

### RESULT 1

T27747  
hypothetical protein ZK1320.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T27747  
R:Berks, M.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: Z20414  
A:Accession: T27747  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-226 <WILL>  
A:Cross-references: EMBL:Z46934; PIDN:CAA87039.1; GSPDB:GNO0020; CESP:ZK1320.1  
A:Experimental source: clone ZK1320  
C:Genetics:  
A:Gene: CESP:ZK1320.1  
A:Map position: 2  
A:Introns: 23/3; 177/3

Query Match 27.7%; Score 328.5; DB 2; Length 226;

Best Local Similarity 33.0%; Pred. No. 5.5e-21;

Matches 73; Conservative 47; Mismatches 94; Indels 7; Gaps 5;

QY 1 MGPLPRTVELFYDVLSPYSWLGFIILCRYQNIW-NINQLRSLITGIMKDSGNKPPGLL 59  
Db 1 MPKLPR-IDFYDVISYPIAFVFOKLEQWKGVTIRYIPFELGAVNMKESGNRPAML 59  
QY 60 PRKGLYNANDKLRLHQLQIPHPKDFLSVMLEKGLSAMRFLTAVNLEHPEMLEKASR 119  
Db 60 PARSIMMTDLKRTAKTFWDIPLTPPLPFMEWIKYRTTGAAMKVLVLVQEDKELMLRAAR 119  
QY 120 ELNWRVMSRNETEPOSILAAAEKAGSAEQAGLLEKIATPKVKNQLKETEAAACRYG 179  
Db 120 EMVRLWSRSEKIFEDQDFVEVLKAGV--KNPEOIVEKSKDEKIKILMENTNKGVDL 177  
QY 180 AFGILP-TTVAHVDQTHMLFGSDRMELLAHLGKWMGP 219  
Db 178 AYGAWINVHTEDGSEHSFSGDRFLHLDLIQPP--QPLP 216

### RESULT 2

T34201  
hypothetical protein D2024.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34201  
R:Du, Z.; Gattung, S.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid D2024.  
A:Reference number: Z21488

30 78.5 6.6 383 2 G71648  
31 78.5 6.6 493 2 T06050  
32 78 6.6 248 2 G59352  
33 78 6.6 313 2 C72784  
34 78 6.6 338 2 A71545  
35 78 6.6 910 2 C69069  
36 78 6.6 1162 2 T40817  
37 78 6.6 1479 2 T17401  
38 77.5 6.5 482 2 T44628  
39 77.5 6.5 634 2 C54454  
40 77 6.5 253 2 G75008  
41 77 6.5 659 2 D72083  
42 76.5 6.5 184 2 T40215  
43 76.5 6.5 497 1 A40487  
44 76.5 6.5 498 1 JT0751  
45 76.5 6.5 843 2 H82362

alanine dehydrogenase (EC 3.2.1.29)  
cellulase (EC 3.2.1.4)  
branched-chain a.a. aminotransferase (EC 2.3.1.42)  
hypothetical prote  
probable C-succinyl-CoA synthetase (EC 6.3.2.1)  
cation-transporter  
botulinum toxin ree  
transcription ree  
probable transposase  
hypothetical prote  
hypothetical prote  
transport ATP band  
hypothetical prote  
ferredoxin--NADP+ reductase (EC 1.6.1.2)  
adenylate cyclase

A:Accession: T34201  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-225 <DUZ>  
A:Cross-references: EMBL:U41011; PIDN:AAA82289.1; CBSP:D2024.7  
C:Genetics:  
A:Gene: CBSP:D2024.7  
A:Introns: 51/1; 177/3

Query Match 25.2%; Score 298.5; DB 2; Length 225;  
Best Local Similarity 33.0%; Pred. No. 2.1e-18;  
Matches 74; Conservative 43; Mismatches 92; Indels 15; Gaps 7;  
QY 3 PLPTVELFYDVLSPYSWLGFEILCRQYQNIWNINLQRLPSLITGIMKDSGNKPPGL--L 59  
DB 2 PNKVKVFFEDVISPSYGFEGITRHSVWKTPIQMKPFFAGVVRHTEN--PGLPLRI 59  
QY 60 PKGLYMANDLKLRLHHLQIPIHFPKDFLSVLMLEKSGLSAMRFLTAVNLEHPEMLEKASR 119  
DB 60 PIKEKYMHDLLFSAQYWGIFPLPKDPTNNMLNTSSIVPQILVASQLRDNVLMEDVAR 119  
QY 120 ELWVRWVRNEDI--TEPQSILAAEAKAGMSAEQAQGLLEKIATPKVNOLKETTEACRY 178  
DB 120 GLWHRFYAYGKPIFTKSQ---VAEVLRLDHLVYDDELVMYMSDSAEVKNILRENTDEAIGN 176  
QY 179 GAFGLP--ITVAHVDGQT-HMFLGSDRMELLAHLLGKMMGPI 218  
DB 177 GCFGAPMWHITDGH--GKVLQTVFGSDRLPQVADFLAEPFGPM 218

RESULT 3  
G83629  
hypothetical protein PA0118 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 15-Sep-2000  
A:Accession: G83629  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950  
A:Accession: G83629  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-195 <STO>  
A:Cross-references: GB:AE004450; GB:AE004091; NID:g9945943; PIDN:AAG03508.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0118

Query Match 16.7%; Score 198; DB 2; Length 195;  
Best Local Similarity 26.1%; Pred. No. 7.2e-10;  
Matches 55; Conservative 41; Mismatches 93; Indels 22; Gaps 5;  
QY 4 LPRTVELFYDVLSPYSWLGFEILCRQYQNIWNINLQRLPSLITGIMKDSGNKPPGLPRKG 63  
DB 1 MSKQIEFFDFGSGTTLTAWTQLPRIAHAHGASIAWRPMLLGGVFKATGNHSPIEVPAK 60  
QY 64 LYMANDLKLRLHHLQIPI---HFPKDFLSVLMLEKSGLSAMRFLTAVNLEHPEMLEKASR 119  
DB 61 RYTLHDLARYAKRYGVPLAFNPAPFINTLTLM-----RGAQGYLGG-----EGFQYILK 109  
QY 120 ELWVRWVRNEDI--TEPQSILAAEAKAGMSAEQAQGLLEKIATPKVNOLKETTEACRY 179  
DB 110 AVEALWVRQNLGKPEVVAQVLAEGAFDPD---FLRLVGDEQVKEGLKATTEAVRRG 166  
QY 180 AFGPLITVAHVDGQTHMFLGSDRMELLAHLL 210  
DB 167 VEGAPSEFF---GDQLFFGQDRLDFVAEVL 193

## RESULT 4

T31286  
2-hydroxychromene-2-carboxylate isomerase homolog - Rhizobium leguminosarum bv. C:Species: Rhizobium leguminosarum bv. viciae  
C:Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 29-Sep-1999  
C:Accession: S72164  
R:Brito, B.; Palacios, J.M.; Ruiz-Argueso, T.; Imperial, J.  
Biochim. Biophys. Acta 1308, 7-11, 1996  
A:Title: Identification of a gene for a chemoreceptor of the methyl-accepting A:Reference number: S72162; MUID:96328256  
A:Accession: S72164  
A:Molecule type: DNA  
A:Residues: 1-194 <BRI>  
A:Cross-references: EMBL:U23040; NID:g780654; PIDN:AAC44312.1; PID:g780657  
A:Experimental source: strain UPM791  
C:Genetics:  
A:Gene: plasmid  
C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase

Query Match 13.9%; Score 164.5; DB 2; Length 194;  
Best Local Similarity 25.2%; Pred. No. 5.4e-07;  
Matches 54; Conservative 34; Mismatches 91; Indels 35; Gaps 8;  
QY 6 RTVELFYDVLSPYSWLGFEILCRQYQNIWNINLQRLPSLITGIMKDSGNKPPGL-LPRKGL 64  
DB 2 RTLDFYDYSYSLALSOVRK---MDVEIAFHPLEIGLDMKQGVNPTTSITCAPKGR 57  
QY 65 YMANDLKLRLHHLQIPIHFPKDFLSV-----MLEKSGLSAMRFLTAVNLEHPEMLEK 116  
DB 58 YVMTDIQWAVHYGVSLNWHPPOLLEIDASRLRLATLVAGOLGAMP--TAV----- 105  
QY 117 ASRELWVRWVRNEDI--TEPQSILAAEAKAGMSAEQAQGLLEKIATPKVNOLKETTEAC 176  
DB 106 --EAIFNIAWISAPLATAAEVAALGAGLDAEE---LAERMDPEAAQDLLEATANAV 160  
QY 177 RYGAFLPITVAHVDGQTHMFLGSDRMELL-AHL 209  
DB 161 SRGVEGAPTLEFV---GDMEFGNDRLHFMQGLH 190

## RESULT 5

T31286  
2-nitrotoluene dioxygenase (EC 1.14.14.1) Rieske iron-sulfur component - Sphingomonas C:Species: Sphingomonas aromaticivorans  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 18-Feb-2000  
C:Accession: T31286  
R:Romane, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Sensen, submitted to the EMBL Data Library, July 1998  
A:Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas A:Reference number: Z20992  
A:Accession: T31286  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-197 <ROM>  
A:Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378427; PIDN:AAD04010.1  
C:Genetics:  
A:Gene: nahD  
A:Gene: plasmid pN1  
C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase  
C:Keywords: oxidoreductase

Query Match 11.8%; Score 140; DB 2; Length 197;  
Best Local Similarity 23.1%; Pred. No. 6.9e-05;  
Matches 50; Conservative 40; Mismatches 94; Indels 32; Gaps 7;  
QY 4 LPRTVELFYDVLSPYSWLGFEILCRQYQNIWNINLQRLPSLITGIMKDSGNKPPG---LLP 60  
DB 1 MTRTHFYDFDISPSYLAQLKLPETARAAAGCTVDYWDIPETPEAKIAAGNYSNREVL 6C  
QY 61 RKGLYMANDLKLRLHHLQIPIHFPKDF-----LSVMLEKSGLSAMRFLTAVNLEHPEML 114

123 MRYWNRNEDITEPOSILAA--AEKAGSAEQAGGLEKIATPKVKNQKLETTAAACRYGA 180  
111 NAVM--GEGIAPDLESALYSEKLGWDSAEHEFLSSNAATE--RYDSQTHAAIERKV 165  
181 FGLPITVAHVGDQTHMLFGSDRMELLALLG 211  
166 FGVP-TWFLGO---EMWGNDRLEFMLESAMG 192

RESULT 8  
H72425  
ABC transporter, ATP-binding protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: H72425  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.;  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richards  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genomic  
A:Reference number: A72200; MUID:99287316  
A:Accession: H72425  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-581 <ARN>  
A:Cross-references: GB:AE001691; GB:AE000512; NID:g4980517; PIDN:AAD35137.1; PID:  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0043  
C:Superfamily: Escherichia coli ABC transporter mdIA; ATP-binding cassette hcmA

Query Match 7.7%; Score 91; DB 2; Length 581;  
Best Local Similarity 19.0%; Pred. No. 4.5;  
Matches 35; Conservative 44; Mismatches 67; Indels 38; Gaps 5;

28 RYONIW-----NINQLRPSLITGIMKDSGNKPPGL-----LPRKGLYMAN 68  
343 REENVWFSDGKNVLDINLDFQPKLYAIVGTGGKSTLMSLNGIYIPKGNIFD 402

69 DLKLLRHHLIQIP---IHFPKDFL-----SYMLEKSGLSAMRFLTAVNLEHP-EMLE 115  
403 EIPLLYNNKLVKQTAANVQDVLFLFSGTILNIRLFDSEIPEERVLEAKRVHLDIIE 462

116 KASRELWVRWNRNEDITEPOSILAAAEKAGM-----SAGQAQGLEKIATPKVKNQK 169  
463 RUPGGVYIEVERGTTLSAGERQLIARAVLFDKIFILDEATSNVDVITETKIOEAL 522

170 ETTE 173  
523 ELSK 526

RESULT 9  
C69127  
flavoprotein A homolog (II) - Methanobacterium thermoautotrophicum (strain Delta H  
C:Species: Methanobacterium thermoautotrophicum  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 24-Sep-1999  
C:Accession: C69127  
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldred,  
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiv  
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N  
J. Bacteriol. 179, 7135-7155, 1997  
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H  
A:Reference number: A69000; MUID:98037514  
A:Accession: C69127  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-410 <MTH>  
A:Cross-references: GB:AE000809; GB:AE000666; NID:g2621265; PIDN:AAB84726.1; PID:  
A:Experimental source: strain Delta H  
C:Genetics:  
A:Gene: MTH220

61 -KIKVAKADLERWAERYGVPLTFEASFACADWNCVLFAREHGKAEAFV-----109  
115 EKASRELWVRWNRNEDITEPOSILAAAEKAGSAEQAGGLEKIATPKVKNQKLETTAA 174  
110 ----DAIRRWGGIDGDRNELAACIAAGL---DPAALIAFVESPAGQNEYRKARSO 161  
175 ACRYGAFGLPITVAHVGDQTHMLFGSDRMELLALLH 210  
162 AIQGVYGAFL--MFVDDQ--IFWGNDRDLFLAEYL 193

RESULT 6  
S17164  
glutathione transferase (EC 2.5.1.18) 13 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-Nov-1993 #sequence\_revision 13-Mar-1997 #text\_change 13-Mar-1997  
C:Accession: S17164  
R:Harris, J.M.; Meyer, D.J.; Coles, B.; Ketterer, B.  
Biochem. J. 278, 137-141, 1991  
A:Title: A novel glutathione transferase (13-13) isolated from the matrix of rat liver  
A:Reference number: S17164; MUID:91354194  
A:Accession: S17164  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-33 <HAR>  
C:Keywords: transferase

Query Match 10.5%; Score 125; DB 2; Length 33;  
Best Local Similarity 71.9%; Pred. No. 0.00013;  
Matches 23; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

3 PLPRTVELFDVLSYMWLGFEILCRYQNIWN 34  
2 PAPRVLEFDVLSYMWLGFEVLRYQHLXN 33

RESULT 7  
C5552  
2-hydroxychromene-2-carboxylate isomerase - Pseudomonas putida plasmid NAH7  
C:Species: Pseudomonas putida  
C:Date: 08-Sep-1995 #sequence\_revision 08-Sep-1995 #text\_change 29-Sep-1999  
C:Accession: C5552  
R:Eaton, R.W. 1994  
J. Bacteriol. 176, 7757-7762, 1994  
A:Title: Organization and evolution of naphthalene catabolic pathways: sequence of the  
from the NAH7 plasmid.  
A:Reference number: A55552; MUID:95095951  
A:Accession: C5552  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-203 <EAT>  
A:Cross-references: GB:U09057; NID:g483790; PIDN:AAA66358.1; PID:g483793  
C:Genetics:  
A:Gene: nahD  
A:Genome: plasmid  
A:Start codon: GTG  
C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase

Query Match 8.0%; Score 95; DB 2; Length 203;  
Best Local Similarity 24.2%; Pred. No. 0.53;  
Matches 51; Conservative 35; Mismatches 97; Indels 28; Gaps 8;

8 VELFDVLSYMWLGFEILCRYQNIWNINLQRLSLITGIMKDSGNKPPGL--LPRKGLY 65  
3 VDFYDFLSPFSYLANQRLSKLAQDYGLTIRYNALDARVTAIGNVGSNRDLKVKLDY 62

66 MANDLKLRLHHLQIPFIHPKDFLSVMLEK-----SLSAMRFLTAVNLEHPMELEKASRELW 122  
63 LKVDLQRAQLYGLIPFPANYSRRMNI GFYSAGAAQAAAYNV-----VF 110

A:Start codon: TTG  
 C:Superfamily: Methanobacterium flavoprotein A  
 C:Keywords: flavoprotein

Query Match 7.6%; Score 90.5; DB 2; Length 410;  
 Best Local Similarity 25.1%; Pred. No. 3.2;  
 Matches 65; Conservative 31; Mismatches 82; Indels 81; Gaps 16;  
 QY 13 DVLSPSWLGFELC-----RYQNIWNINLQRLPSLTIGIMKDSGNK----- 54  
 DB 99 EVLKRY---GSEICTAKAAGLRQHSIPQDTPMQ---TVRTGSDTDLGGKTLTFLEAP 152  
 QY 55 -----PPG---LLPRKGLYMANDLLRHLHLPPIHPKDFLSVMLEKGSLSAMRF---L 103  
 DB 153 MLHWPDSDMTLLEEGILFSND---AFGQHLCSIKRFDKDVPEAVLMD---AAKKFYANLL 207  
 QY 104 TAVNLEHPEMLERAS--RELWM-----RVWSRNEDEITEPQSILAAAEKAGMSA 149  
 DB 208 TPLS---PLVIRKFEVKEGLLEKIGMIAPSHQGIW-----TEPLKIIAA-----YT 252  
 QY 150 EQAQGLLEKIATPKVNQLKETTAAACRYGAFGLPITVAHVGDQTHMLFGSDRMELLAHL 209  
 DB 253 DWATGCRDKAT--IYDTMHSYEMLAHMAEGL--MAADVDSMHFLHEDERSEIVKNI 309  
 QY 210 LGEK-----WMGPPIP 219  
 DB 310 LESKAVFGSPTMFNGPFP 328

RESULT 10  
 SNECPI  
 N:pitilysin (EC 3.4.24.55) precursor [validated] - Escherichia coli  
 C:Species: Escherichia coli  
 C:Date: 31-Mar-1993 #sequence\_revision 31-Oct-1997 #text\_change 18-Feb-2000  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617  
 A:Accession: F65064  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-962 <BLAT>  
 A:Cross-references: GB:A5000365; GB:U000096; NID:G2367163; PIDN:AAC75860.1; PID:G2367164;  
 A:Experimental source: Strain K-12, substrain MG1655  
 R:Claverie-Martin, F.; Diaz-Torres, M.R.; Kushner, S.R.  
 Gene 54, 185-195, 1987  
 A:Title: Analysis of the regulatory region of the protease III (ptr) gene of Escherichia coli  
 A:Reference number: A29093; MUID:88005781  
 A:Accession: A29093  
 A:Molecule type: DNA  
 A:Residues: 1-276; 'HYHSLR', 283, 'W', 285-296 <CLA>  
 A:Cross-references: GB:M17095; NID:G147390; PIDN:AAA24436.1; PID:G147391  
 A:Experimental source: strain K12  
 R:Finch, P.W.; Wilson, R.E.; Brown, K.; Hickson, I.D.; Emmerson, P.T.  
 Nucleic Acids Res. 14, 7695-7703, 1986  
 A:Title: Complete nucleotide sequence of the Escherichia coli ptr gene encoding protease III  
 A:Reference number: A25765; MUID:87040734  
 A:Accession: A25765  
 A:Molecule type: DNA  
 A:Residues: 1-962 <FTN>  
 A:Cross-references: GB:X06227; NID:G42560; PIDN:CAA29576.1; PID:G42561  
 R:Becker, A.B.; Roth, R.A.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 3835-3839, 1992  
 A:Title: An unusual active site identified in a family of zinc metalloendopeptidases.  
 A:Reference number: A38854; MUID:92237263  
 C:Contents: annotation; active site  
 C:Genetics: ptr  
 A:Gene: ptr

A:Map position: 61

C:Function:

A:Description: endopeptidase degrades small peptides [validated; MUID:92237263]  
 A:Pathway: protein degradation  
 C:Superfamily: insulysin  
 C:Keywords: hydrolase; metalloproteinase; monomer; periplasmic space; protein de  
 F:1-23/Domain: signal sequence #status predicted <Sig>  
 F:24-962/Product: pitrilysin #status experimental <MAT>  
 F:88,92/Binding site: zinc (His) #status experimental  
 F:91/Active site: Glu #status experimental

Query Match 7.6%; Score 89.5; DB 1; Length 962;  
 Best Local Similarity 26.3%; Pred. No. 12;  
 Matches 30; Conservative 13; Mismatches 48; Indels 23; Gaps 3;  
 QY 126 WSRNEDITEPQSILAAAEKAGMSAEQA-----OGLEKIATPKVNQ 167  
 DB 731 WCRNKDVVYDKQSVIFERAGNSTDSALAAVFVPTGYDEYTSAYSSLLGQIVQPFYQ 790  
 QY 168 LKETTEACRYGAFGLPITVAHVGDQTHMLFGSDRMELLAHLGKMGPIPPA 221  
 DB 791 LR--TEEQGLYAVFAFPMSYGVGRQMGFLQSDNQ----PSFLWERYKAFFPTA 839

RESULT 11  
 I49343  
 Probable isomerase doxJ - Pseudomonas sp. (strain C18)  
 C:Species: Pseudomonas sp.  
 C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 18-Sep-1998  
 R:Denome, S.A.; Stanley, D.C.; Olson, E.S.; Young, K.D.  
 J. Bacteriol. 175, 6890-6901, 1993  
 A:Title: Metabolism of dibenzothiophene and naphthalene in Pseudomonas strains:  
 A:Reference number: A49343; MUID:94042852  
 A:Accession: I49343  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-199 <DEN>  
 A:Cross-references: GB:M60405  
 C:Genetics: doxJ  
 C:Superfamily: 2-hydroxychromene-2-carboxylate isomerase

Query Match 7.5%; Score 89; DB 2; Length 199;  
 Best Local Similarity 24.2%; Pred. No. 1.7;  
 Matches 51; Conservative 37; Mismatches 95; Indels 28; Gaps 9;  
 QY 8 VELFYDVLSPSWLGFELCRYQNIWNINLQRLPSLTIGIMKDSGNKPPGL--LPRKGLY 65  
 DB 3 VDFYDFELSPSYLANHRLSKLAQDYGFISIRYAIQDLARVKIAIGNVGPSNRDLIVKLDY 62  
 QY 66 MANDLKLLRHLHLPPIHPKDFLSVMLEKGSLS--SAMRELTAVNLEHPEMLEKASRELW 122  
 DB 63 LKVDLQRAEYIEPLVPFANYNSRRMTGLYSGAMAAQTGAYVNV-----VF 110  
 QY 123 MRVWSRNEDEITEPQSILAA--AEKAGMSAEQAQGLEKIATPKVNQLKETTEACRYGA 180  
 DB 111 NAVW--CGDIAPDLESPLVSEKLGWDSAFE---DFISSDAATERYDEQTHAIERKV 155  
 QY 181 FGLPITVAHVGDQTHMLFGSDRMELLAHLG 211  
 DB 166 FGVP--TMFLGJ---EMWNGNDRFLFLENVAVG 192

RESULT 12  
 T51712

threonine dehydratase/deaminase [Imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000  
 C:Accession: T51712  
 R:Mourad, G.S.; Emerick, R.M.; Marion, A.L.; Smith, A.M.





Thu Feb 15 15:59:14 2001

us-08-978-174-1.rpr

Page 6

Search completed: February 15, 2001, 15:52:42  
Job time: 114 sec

Result No.	Query No.	Score	Match	Length	DB	ID	Description
1	1	869	73.3	225	1	GTK1_RAT	P24473 rattus norv
2	2	328.5	27.7	226	1	YS21_CAEEL	Q09652 caenorhabdi
3	3	298.5	25.2	225	1	Y157_CAEEL	Q18973 caenorhabdi
4	4	95	8.0	203	1	N4HD_PSEPU	Q51948 pseudomonas
5	5	89.5	7.6	962	1	N4RD_ECOLI	P05458 escherichia
6	6	89	7.5	212	1	SHAD_PSESP	Q52462 pseudomonas
7	7	88.5	7.5	613	1	SG2_BOVIN	P20616 bos taurus
8	8	84.5	7.1	316	1	M1AA_TREPA	Q83644 treponema p
9	9	84.5	7.1	971	1	YMRP_YEAST	P49955 saccharomyc
10	10	83.5	7.0	488	1	CATA_LISSE	P24188 listeria se
11	11	83.5	7.0	700	1	SC2_DROME	P20480 drosophila
12	12	81.5	6.9	619	1	SG2_RAT	P10362 rattus norv
13	13	81.5	6.9	1061	1	DPOL_ADE12	P06538 human adeno
14	14	81	6.8	274	1	YG73_SYNY3	P74261 synecocyst
15	15	80.5	6.8	339	1	PURK_BRUME	P52559 brucella me
16	16	80.5	6.8	705	1	CS5_YEAST	P25862 saccharomyc
17	17	80.5	6.8	1073	1	HSR_PIG	P35204 sus scrofa
18	18	80	6.8	540	1	TOP1_AQUAE	O66893 aquifex aeo
19	19	80	6.8	1162	1	XBEN_CLOBO	P46082 clostridium
20	20	79.5	6.7	547	1	CATX_BACSU	P94377 bacillus su
21	21	79.5	6.7	1073	1	HSRP_HUMAN	P25092 homo sapien
22	22	78.5	6.6	383	1	PNAA_RCPR	P41077 rickettsia
23	23	78	6.6	1162	1	XCEN_CLOBO	Q06366 clostridium
24	24	77.5	6.5	634	1	YB36_MFTJA	Q58633 methanococ
25	25	77	6.5	494	1	ADRO_RAT	P56522 rattus norv
26	26	76.5	6.5	962	1	YBX7_SCHPO	Q10201 schizosacch
27	27	76.5	6.5	1056	1	DPOL_ADE02	Q10261 human adeno
28	28	76.5	6.5	1193	1	DPOL_ADE04	P87503 human adeno
29	29	76	6.4	406	1	PDK3_HUMAN	Q5120 homo sapien
30	30	76	6.4	646	1	POPD_RHIS3	P72339 r nodg bifu
31	31	76	6.4	2208	1	POLN_MANCV	Q69014 manchester
32	32	75.5	6.4	744	1	GYRA_AQUAE	O67108 aquifex aeo
33	33	75.5	6.4	887	1	GLND_KLEPN	P41393 klebsiella

Best Local Similarity 69.3%; Pred. No. 1.5e-68;  
Matches 156; Conservative 38; Mismatches 31; Indels 0; Gaps 0;

QY 2 GPLPRTVELFDVLSYSLGFEILCRYQNIWNTNINLQRLPSLITGIMKDSGNKPPGLPR 61  
DB 1 GPARVLEFVLSYSLGFEVLCRYQHLWNKILKRLPALLAGIMKDSGNQPPAMVPH 60

QY 62 KGLYMANDLKLRLHHLQIPIHFPKDFLSVLMLEKGLSAMRFLTAVNLEHPMELEKASREL 121  
DB 61 KGQYLKEIPLKQLFQVPMSPKDFEGEHVKKGTVNARFLTAVSMQPEMLEKYSREL 120

QY 122 WMRVWSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLTETTAACRYGAF 181  
DB 121 WMRVWSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLTETTAACRYGAF 180

QY 182 GLPITVAHVGDGTHMLFGSDRMELLALHLEKMGKMPPIPPAVNARL 226  
DB 181 GLPITVAHVGDGTHMLFGSDRMELLALHLEKMGKMPPIPPAVNARL 225

RESULT 2  
ID YS21\_CAEEL STANDARD; PRT; 226 AA.  
AC Q09652;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DE HYPOTHETICAL 26.5 KDA PROTEIN ZK1320.1 IN CHROMOSOME II.  
GN ZK1320.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Berks M.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. KAPPA FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: Z46934; CAA87039.1;  
DR WORMPEP: ZK1320.1; CE01698.  
KW Hypothetical protein.  
SQ SEQUENCE 226 AA; 26482 MW; DBA39A5994300164 CRC64;

Query Match 27.7%; Score 328.5; DB 1; Length 226;  
Best Local Similarity 33.0%; Pred. No. 1.3e-21;  
Matches 73; Conservative 47; Mismatches 94; Indels 7; Gaps 5;

QY 1 MGPLPRTVELFDVLSYSLGFEILCRYQNIW-NINLQRLPSLITGIMKDSGNKPPGLL 59  
DB 1 MPKLPRI-IDFVDFVISPYIAFEVFKLETKQKGVIRIYIPFLGAVMKESGNRPPL 59

QY 60 PRGLYVANDLKLRLHHLQIPIHFPKDFLSVLMLEKGLSAMRFLTAVNLEHPMELEKASR 119  
DB 60 PARSIMMTDLKRTAKFDWIDPLTPPLFMWIKKYRTTGAMKVLVLLQEQDKELMLRAAR 119

QY 120 ELWVRVWSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLTETTAACRYG 179  
DB 120 EMVRLWSRSRKIEDQDFEVLKAVGV--KNPEQIVKESKDEKVIKILMTNKGVDLM 177

QY 180 AFGLP-ITVAHVGDGTHMLFGSDRMELLALHLEKMGKMPPI 219  
DB 178 AYGAPINWHTEDGSEHFFGSDRFLIADLLQOP--QPLP 216

RESULT 3  
ID YS7\_CAEEL STANDARD; PRT; 225 AA.  
AC Q18973;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HYPOTHETICAL 26.0 KDA PROTEIN D2024.7 IN CHROMOSOME IV.  
GN D2024.7.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Du Z., Gatung S.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. KAPPA FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U41011; AAA82289.1;  
DR WORMPEP: D2024.7; CE04296.  
KW Hypothetical protein.  
SQ SEQUENCE 225 AA; 2300A3D6762B7232 CRC64;

Query Match 25.2%; Score 298.5; DB 1; Length 225;  
Best Local Similarity 33.0%; Pred. No. 5.2e-19;  
Matches 74; Conservative 43; Mismatches 92; Indels 15; Gaps 7;

QY 3 PLPRTVELFDVLSYSLGFEILCRYQNIWNTNINLQRLPSLITGIMKDSGNKPPGL--L 59  
DB 2 PNKRVKFEFDVLSYSLGFEIGTRHSVWKTPIQMKPFFAGVVRHTEN--FGLPLRI 59

QY 60 PRGLYVANDLKLRLHHLQIPIHFPKDFLSVLMLEKGLSAMRFLTAVNLEHPMELEKASR 119  
DB 60 PIKEYTHMKDLFLSAQWGPFLPKDYTNMMLNTSSIPQRLVASQLRDNVLMEDVAR 119

QY 120 ELWVRVWSRNEDITEPOSILAAAEKAGMSAEQAQGLLEKIATPKVKNQKLTETTAACRY 178  
DB 120 GLWHRFYAYGKPIFTKSQ---VAEVLRLDHVKDVEDVLMVMSDSAEVKNILRENTDEIGN 176

QY 179 GAFGLP---ITVAHVGDGTHMLFGSDRMELLALHLEKMGKMPPI 218  
DB 177 GCGFAPWMHTDGH--GKVLQTVFGSDRLPOVADFLAEPEFKGPM 218

RESULT 4  
ID NAHD\_PSEPU STANDARD; PRT; 203 AA.  
AC Q51948;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE 2-HYDROXYCHROMENE-2-CARBOXYLATE ISOMERASE (HCCA ISOMERASE).  
GN NAHD.  
OS Pseudomonas putida.  
OC Plasmid NAH7.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-G7 / ATCC 17485;  
RX MEDLINE-95095951; PubMed-8002605;  
RA Eaton R.W.;



DR PROSITE; PS00143; INSULINASE; 1.  
 KW Hydrolase; Metalloprotease; Magnesium; Periplasmic; Zinc; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 962  
 FT METAL 88 88  
 FT ACT\_SITE 91 91  
 FT METAL 92 92  
 FT METAL 169 169  
 FT METAL 169 169  
 FT MUTAGEN 88 88  
 FT MUTAGEN 91 91  
 FT MUTAGEN 92 92  
 FT MUTAGEN 162 162  
 FT MUTAGEN 169 169  
 FT MUTAGEN 204 204  
 FT CONFLICT 277 284  
 SQ SEQUENCE 962 AA; 107708 MW; 0558682C2F1A0540 CRC64;

Query Match 7.6%; Score 89.5; DB 1; Length 962;  
 Best Local Similarity 26.3%; Pred. No. 4.4;  
 Matches 30; Conservative 13; Mismatches 48; Indels 23; Gaps 3;

QY 126 WSRNEDITEPQSILAAAEKAGMSAEQA-----OGLEKIATPKVKNQ 167  
 DB 731 WCRKDVVDKOSVIFKAGNSTDSALAAVFTGYDEYTSAYSSLLGQIVQPFYQ 790  
 QY 168 LKETTEAACRYGAFGLPITVAHVGDQTHMFLGSDRMELLALLHLLGKWMGPPIPA 221  
 DB 791 LR--TEEQLYGAVFAFPMISVGRQWGMGFLLOSNDKQ---PSFLWERYKAFPPFA 839

RESULT 6  
 NAHD\_PSESP STANDARD; PRT; 212 AA.  
 AC Q52462;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE 2-HYDROXYCHROMENE-2-CARBOXYLATE ISOMERASE (HCCA ISOMERASE).  
 GN DOXJ.  
 OS Pseudomonas sp. (strain C18).  
 OC Plasmid.  
 OC Bacteria; Proteobacteria.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-94042852; PubMed-8226631;  
 RA Denome S.A., Stanley D.C., Olson E.S., Young K.D.;  
 RT "Metabolism of dibenzothioephene and naphthalene in Pseudomonas strains: complete DNA sequence of an upper naphthalene catabolic pathway";  
 RT J. Bacteriol. 175:6890-6901(1993).  
 CC -1- FUNCTION: CATALYZES THE ISOMERIZATION OF 2-HYDROXYCHROMENE-2-CARBOXYLATE (HCCA) TO TRANS-O-HYDROXYBENZYLIDENEPYRUVATE (THBPA).  
 CC THE OPTIMUM PH FOR THE ENZYME IS 10. THE REACTION IS REVERSIBLE.  
 CC -1- PATHWAY: UPPER NAPHTHALENE CATABOLIC PATHWAY WHICH INVOLVES CONVERSION OF NAPHTHALENE TO SALICYLATE. AND CATABOLISM OF DIBENZOTHIOEPHENE (DBT) AND PHENANTHRENE. CONVERTS PHENANTHRENE TO 1-HYDROXY-2-NAPHTHOIC ACID AND THE METABOLISM OF DBT IS LIMITED TO OXIDATION OF THE AROMATIC RING.  
 CC -1- MISCELLANEOUS: DOXH AND DOXJ ENCODE DIFFERENT ENZYMES THAT MAY HAVE INTERCHANGEABLE FUNCTIONS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; M60405; AAA16133.1;  
 DR Isomerase; Plasmid; Aromatic hydrocarbons catabolism.  
 KW Isomerase; Plasmid; Aromatic hydrocarbons catabolism.  
 SQ SEQUENCE 212 AA; 24039 MW; 5EF96A619913DB4F CRC64;

Query Match 7.5%; Score 89; DB 1; Length 212;  
 Best Local Similarity 24.2%; Pred. No. 0.75;  
 Matches 51; Conservative 37; Mismatches 95; Indels 28; Gaps 9;

QY 8 VELYDVLPSPSWLGFELCRYQNIWNINLQRLPSLIITGIMKDSGNKPPGL--LPRKGLY 65  
 DB 16 VDFYDFLSPVSYLANHRLSKLAQDYGFPSIRYAYDLARVKAIGNVGPNSNRDLIVKLDY 75  
 QY 66 MANDLKLRLHILPIHFPKDFLSVMLEKSL--SAMREFTA-VNLEHPEMLEKASRELW 122  
 DB 76 LKVDLQWRWALYELPLVFPANYNRRMTGLYSGAMAQTGAVNV-----VF 123  
 QY 123 MRYWSNEDITEPQSILAA--AEKAGMSAEQAQGLEKATKPKVKQLKETTAAACRYGA 180  
 DB 124 NAVW--GDGIA/PDLES/PALVSEKLGWDRSAFE---DFISSAATERYDEQTHAAIERKV 178  
 QY 181 FGLPITVAHVGDQTHMFLGSDRMELLALLHLLG 211  
 DB 179 FGVP-TMFLGD---EMWGNDRFLMENAUG 205

RESULT 7  
 SG2\_BOVIN STANDARD; PRT; 613 AA.  
 AC P20616;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE SECRETOGGRANIN II PRECURSOR (SGII) (CHROMOGRANIN C).  
 GN SCG2 OR CHGC.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-90264409; PubMed-2345170;  
 RA Fischer-Colbrie R., Gutierrez J., Hsu C.M., Tacangelo A., Eiden L.E.;  
 RT "Sequence analysis, tissue distribution and regulation by cell depolarization, and second messengers of bovine secretogranin II (chromogranin C) mRNA";  
 RT J. Biol. Chem. 265:9208-9213(1990).  
 CC -1- FUNCTION: SECRETOGGRANIN II IS A NEUROENDOCRINE SECRETORY GRANULE PROTEIN, WHICH IS THE PRECURSOR FOR BIOLOGICALLY ACTIVE PEPTIDES.  
 CC -1- SUBCELLULAR LOCATION: NEUROENDOCRINE AND ENDOCRINE SECRETORY GRANULES.  
 CC -1- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.  
 CC -1- SIMILARITY: BELONGS TO THE CHROMOGRANIN / SECRETOGGRANIN PROTEIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; J05468; AAA30760.1;  
 DR PIR; A35296; A35296.  
 DR INTERPRO; IPR001990;  
 DR PFAM; PF01271; Granin; 1.  
 DR PROSITE; PS00422; GRANINS\_1; 1.  
 KW Sulfatation; Cleavage on pair of basic residues; Calcium-binding; Signal.  
 FT SIGNAL 1 27 POTENTIAL.  
 FT PROPEP 28 30 POTENTIAL.  
 FT CHAIN 31 613 SECRETOGGRANIN II.  
 FT PEPTIDE 181 213 UNKNOWN ACTIVITY PEPTIDE (PROBABLE).  
 FT MOD\_RES 150 150 SULFATATION (BY SIMILARITY).  
 SQ SEQUENCE 613 AA; 70356 MW; 5DC079F59D83516 CRC64;

Qy	13	DVLSVSWLGEFELCRYQNIWNLQRLPSLITGIMKDSGNKPPGLPRKGLYMANDLKL	72
Db	80	DVCDPTE--EYNVFRQAVYGI---VPSIL-----RAHKVPIVGGTGLYLD---AV	124
Qy	73	LRHHLQIPIHPKDFLSVMLEKGSLSAM-----REFLAVNL-	108
Db	125	LRQYALVPVE-RNQAALRASLRGASLSHMRVYFSLKDSHAVHNKNTDLEDPARLMRAIEIA	183
Qy	109	-----EHPMELEKASRELWMRVWSNEDITEPQSIILAAAEKAGMSAEQAQGLLEKIA	160
Db	184	VFHATHPELLOQA-RETRPMRAKVYGIQYPRSMRLRARIRARLEQRIGRLIEEVA	238
RESULT 9			
YMBP_YEAST	STANDARD;	PRT;	971 AA.
ID	YMBP_YEAST		
AC	P49955;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	15-JUL-1998 (Rel. 36, Last annotation update)		
DE	HYPOTHETICAL 110.0 KDA PROTEIN IN MSU1-HAS1 INTERGENIC REGION.		
GN	YMR288W OR YMR021.14.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;		
OC	Saccharomycetaceae; Saccharomycetes.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-S288C / AB972;		
RA	Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;		
RL	Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.		
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).		
CC	-1- SIMILARITY: STRONG, TO S.POMBE SPAC2F1.09C.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; 249704; CAA89786.1;		
DR	SGD; S0004901; YMR288W.		
KW	Hypothetical protein; Transmembrane.		
FT	TRANSMEM 232 252	POTENTIAL.	
FT	TRANSMEM 300 320	POTENTIAL.	
FT	TRANSMEM 336 356	POTENTIAL.	
FT	TRANSMEM 643 663	POTENTIAL.	
FT	TRANSMEM 748 768	POTENTIAL.	
FT	TRANSMEM 782 802	POTENTIAL.	
FT	TRANSMEM 877 897	POTENTIAL.	
FT	TRANSMEM 905 925	POTENTIAL.	
FT	SEQUENCE 971 AA; 110027 MW; 27D26E4252A788E2 CRC64;		
Query Match 7.1%; Score 84.5; DB 1; Length 971;			
Best Local Similarity 21.7%; Pred. No. 12;			
Matches 36; Conservative 31; Mismatches 64; Indels 35; Gaps 7;			
Qy	5	PRTVELYDVLSVSWLGEFELCRYQNIWNLQRLPSLITGIMKDSGNKPPGLPR-KG	63
Db	387	PYGIEVFNVLPEP-----LWKGIRSHRGKVLSSFLSKAVGSMPLMDPEYAG	432
Qy	64	LYMANDLKLRLHHLQIPIHPKDFLSVMLEKGSLSAMREFLAVNLHPEML- EKASRELW	122
Db	433	YTTTEAMRIIRREFDSPDEMKTKTLLVLOK--CSAVESIT-----PKFLREIEAIEFF	484
Qy	123	MRVWSNEDITEP-----QSTILAAAEKAGMSAEQAQGLLEKATP	162
Db	485	QKFWRVYALDRPLNKVVYTTVTLAKLGCYSY-----TIDKLLTP	525
RESULT 10			

Qy	6	RVVLFYDVLSVSWLGEFELCRYQNIWNLQRLPSLITGIMKDSGNKPPGLLP	60
Db	333	RAIRLEFKPLDP-----QSIYQLIEISRNLIQIPPEDLIDMLK-TGEKP--VEP	377
Qy	61	RKGLYMANDLKLRLHHLQIPIHPKDFLSVMLEKGSLSAMREFLAVNLHPEMLEKASRE	120
Db	378	EQ-----ELEIPEV-PED-----ISEVDLQHPDLFQN----	403
Qy	121	LMRVWSNEDITEPQSIILAAAEKAGMSAEQAQGLL--EKIATPK	163
Db	404	---KMLSKNGYKAPAGHAVAALPEGLSVEDILNLGMSAANPK	445
RESULT 8			
YMBP_YEAST	STANDARD;	PRT;	316 AA.
ID	MIAA_TREPA		
AC	O83644;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	TRNA DELTA(2)-ISOPENTENYLPIROPHOSPHATE TRANSFERASE (EC 2.5.1.18) (IPP		
DE	TRANSFERASE).		
GN	MIAA OR TP0637.		
OS	Treponema pallidum.		
OC	Bacteria; Spirochaetales; Spirochaetaceae; Treponema.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPRAINE-NICHOLS;		
RA	Medline-98332770; PubMed-9665876;		
RA	Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,		
RA	Dodson K., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,		
RA	Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,		
RA	Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,		
RA	McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,		
RA	Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,		
RA	Venter J.C.;		
RA	"Complete genome sequence of Treponema pallidum, the syphilis		
RT	spirochete."		
RL	Science 281:375-388(1998).		
CC	-1- FUNCTION: CATALYZES THE FIRST STEP IN THE BIOSYNTHESIS OF		
CC	2-METHYLTHIO-N6-(DELTA(2)-ISOPENTENYL)-ADENOSINE (MS[21][61A])		
CC	ADJACENT TO THE ANTICODON OF SEVERAL TRNA SPECIES (BY SIMILARITY).		
CC	-1- CATALYTIC ACTIVITY: ISOPENTENYL DIPHOSPHATE + TRNA -		
CC	PIROPHOSPHATE + TRNA CONTAINING 6-ISOPENTENYLADENOSINE.		
CC	-1- SIMILARITY: BELONGS TO THE IPP TRANSFERASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; AE001238; AAC65611.1;		
DR	TIGR; TP0637;		
DR	INTERPRO: IPR002627;		
DR	PFAM; PF01715; IPT: 1.		
KW	Transferase; Nucleotidyltransferase; trna processing; ATP-binding.		
NP_BIND	27 34	ATP (POTENTIAL).	
FT	SEQUENCE 316 AA; 36174 MW; E4BE144E56ED5208 CRC64;		
Query Match 7.1%; Score 84.5; DB 1; Length 316;			
Best Local Similarity 23.3%; Pred. No. 3;			
Matches 41; Conservative 30; Mismatches 60; Indels 45; Gaps 8;			







Best Local Similarity 22.4%; Pred. No. 9.8; Matches 52; Conservative 37; Mismatches 70; Indels 73; Gaps 12;

QY 1 MGPIPRVVELFYDLSPYSWLG--FEILCRQNIWNINLQRLPSLTIGMKD-----SG 52  
DB 451 VGVIPRTVDLFDLSIRYNGWYEIKATFLEIYN---EVLVDLISNEQKDMIRMAKN 507

QY 53 NKPPGLPRKGLYMAN---DLKLLRHHLQIPIHFPKDFLSVLMLEKSLSRMFLITAVNLE 109  
DB 508 NK-----NIIYVSNITEETVLDPNHLRLHMTAK-----MNRATAS-----TAGN-- 547

QY 110 HPEMLEKASRELWVRWNRNEDITEPQSILAAAEKAGMS-----AEQAGGLEK 158  
DB 548 -----ERS-----SRSHAVTKLEIGRHAERKOEISVGSINLVDLASESPKSTRM 593

QY 159 IATPKVKNQKTEACRYGAFGLPITVAHVQDQTHMLFGSDRMELLALL 210  
DB 594 TETKNINRSLSELTN-----VILALQKQDHIPY---RNSKLTHLL 631

RESULT 12  
SG2\_RAT  
ID SG2\_RAT STANDARD; PRT; 619 AA.  
AC P10362;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE SECRETOGRANIN II PRECURSOR (SGII) (CHROMOGRANIN C).  
GN SCG2 OR CHGC OR SCG-2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-89098327; PubMed-3211750;  
RA Gerdes H.H., Philipps E., Huttner W.B.;  
RT "The primary structure of rat secretogranin II deduced from a cDNA sequence.";  
RL Nucleic Acids Res. 16:11811-11811(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-93309708; PubMed-8321414;  
RA Kakar S., Wei N., Mulchahey J.J., Lebeuf R.D., Neill J.D.;  
RT "Regulation of expression of secretogranin II mRNA in female rat pituitary and hypothalamus.";  
RL Neuroendocrinology 57:422-431(1993).  
CC -!- FUNCTION: SECRETOGRANIN II IS A NEUROENDOCRINE SECRETORY GRANULE PROTEIN, WHICH IS THE PRECURSOR FOR BIOLOGICALLY ACTIVE PEPTIDES.  
CC -!- SUBCELLULAR LOCATION: NEUROENDOCRINE AND ENDOCRINE SECRETORY GRANULES.  
CC -!- MISCELLANEOUS: BINDS CALCIUM WITH A LOW-AFFINITY.  
CC -!- SIMILARITY: BELONGS TO THE CHROMOGRANIN / SECRETOGRANIN PROTEIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X13618; CAA31950.1; -  
CC EMBL: M93669; AAM42135.1; -  
CC PIR: S02180; S02180.  
CC INTERPRO: IPR001990; -  
CC PFAM: PF01271; granin; 1.  
CC PROSITE: PS00422; GRANINS\_1; 1.  
CC Sulfatation: Cleavage on pair of basic residues; Calcium-binding; Signal.  
CC  
CC KW SIGNAL 1 30  
CC FT CHAIN 31 619  
CC FT PEPTIDE 184 216  
CC FT  
CC SECRETOGANIN II.  
CC UNKNOWN ACTIVITY PEPTIDE (PROBABLE).

RA MEDLINE-91043032; PubMed-2146510;  
RA Walker R.A., Salmon E.D., Endow S.A.;  
RT "The Drosophila claret segregation protein is a minus-end directed motor molecule.";  
RL Nature 347:780-782(1990).  
RN [5]  
RP CHARACTERIZATION.  
RX MEDLINE-94155838; PubMed-8112290;  
RA Lockhart A., Cross R.A.;  
RT "Origins of reversed directionality in the ncd molecular motor.";  
REMBO J. 13:751-757(1994).  
RN [6]  
RN MUTANT ALLELE NCD(D).  
RX MEDLINE-91122049; PubMed-1825056;  
RA Komma D.J., Horne A.S., Endow S.A.;  
RT "Separation of meiotic and mitotic effects of claret non-disjunctional on chromosome segregation in Drosophila.";  
REMBO J. 10:419-424(1991).  
RN [7]  
RN CHARACTERIZATION OF MUTANT ALLELE NCD(D).  
RX MEDLINE-96283629; PubMed-8670831;  
RA Moore J.D., Song H., Endow S.A.;  
RT "A point mutation in the microtubule binding region of the Ncd motor protein reduces motor velocity.";  
REMBO J. 15:3306-3314(1996).  
RN [8]  
RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 335-700.  
RX MEDLINE-96195067; PubMed-8606780;  
RA Sablin E.P., Kull F.J., Cooke R., Vale R.D., Fletcher R.J.;  
RT "Crystal structure of the motor domain of the kinesin-related motor ncd.";  
RL Nature 380:555-559(1996).  
CC -!- FUNCTION: NCD IS REQUIRED FOR NORMAL CHROMOSOMAL SEGREGATION IN MEIOSIS, IN FEMALES, AND IN EARLY MITOTIC DIVISIONS OF THE EMBRYO. THE NCD MOTOR ACTIVITY IS DIRECTED TOWARD THE MICROTUBULE'S MINUS END.  
CC -!- MISCELLANEOUS: NCD(D) IS A MUTANT ALLELE THAT SHOWS ABNORMAL CHROMOSOMAL SEGREGATION.  
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. NCD SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X52814; CAA36998.1; -  
CC EMBL: M33932; AAM28716.1; -  
CC EMBL: AE003771; AAF56942.1; -  
CC EMBL: X57475; CAA40713.1; -  
CC PIR: A35624; A35624.  
CC PIR: S09748; S09748.  
CC HSP: P17119; 3KAR.  
CC FLXBASE: FBgn0002924; ncd.  
CC INTERPRO: IPR001732; -  
CC PFAM: PF00225; kinesin; 1.  
CC PRINTS: PR00380; KINESINHEAVY.  
CC PROSITE: PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
CC PROSITE: PS00067; KINESIN\_MOTOR\_DOMAIN2; 1.  
CC Motor protein: Cell division; Microtubules; ATP-binding; Coiled coil;  
CC  
CC Meiosis; Mitosis. 346  
CC DOMAIN 196 700  
CC COILED COIL (POTENTIAL).  
CC MECHANOCHEMICAL (MOTOR) (BY SIMILARITY).  
CC FT DOMAIN 347 700  
CC FT NP\_BIND 434 441  
CC FT ATP (BY SIMILARITY).  
CC FT MUTAGEN 556 556  
CC V->F: IN NCD(D); REDUCES MOTOR VELOCITY.  
CC FT CONFLICT 697 697  
CC S -> N (IN REF. 1).  
CC SEQUENCE 700 AA; ADE043CBCE7FD561 CRG64;  
CC  
CC Query Match 7.0%; Score 83.5; DB 1; Length 700;



QY 67 ANDL-----KLRHHIQIPIHPK---DFLSVMLEKGSLSAMRFLTAVNLEHPMELEKA 117  
Db 152 TADCVDPSTPKVLRSSAGSSLLPQQLQSLPPLLEKFTHTQGLQIATV-----PQA 203  
QY 118 SRELWVRWNRNEDITEPOSILAAAKAGMAEQAQGLLEKIATPK 163  
Db 204 TQTLW-----EIDFQRTIVIFGSGQGLSAPVLELTHQVAIPQ 243

RESULT 15

PURK\_BRUME STANDARD; PRT; 339 AA.  
AC P52559;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE PHOSPHORIBOSYLAMINOIMIDAZOLE CARBOXYLASE ATPASE SUBUNIT (EC 4.1.1.21)  
DE (AIR CARBOXYLASE) (AIRC).  
GN PURK.  
OS Brucella melitensis.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Brucellaceae; Brucella.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-16M;  
RA Warren R., Hoover D., Hadfield T., Drazek S.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: POSSESSES AN ATPASE ACTIVITY THAT IS DEPENDENT ON THE  
CC PRESENCE OF AIR (AMINOIMIDAZOLE RIBONUCLEOTIDE). THE ASSOCIATION  
CC OF PURK AND PURE PRODUCES AN ENZYME COMPLEX CAPABLE OF CONVERTING  
CC AIR TO CAIR EFFICIENTLY UNDER PHYSIOLOGICAL CONDITION  
CC (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: 1-(5-PHOSPHORIBOSYL)-5-AMINO-4-IMIDAZOLE-  
CC CARBOXYLATE -> 1-(5-PHOSPHORIBOSYL)-5-AMINOIMIDAZOLE + CO(2).  
CC -1- PATHWAY: SIXTH STEP IN DE NOVO PURINE BIOSYNTHESIS.  
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE PURK / PURT FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U10241; AAA57003.1;  
KW Purine biosynthesis; Lyase; Decarboxylase.  
SQ SEQUENCE 339 AA; 37166 MW; 951B0DC5E637324E CRC64;

Query Match 6.8%; Score 80.5; DB 1; Length 339;  
Best Local Similarity 21.7%; Pred. No. 7.3;  
Matches 65; Conservative 25; Mismatches 96; Indels 113; Gaps 14;  
QY 3 PLPRIVELYDVLS---PYSWLGFEILCRQNIWNIQLRPSLITGIMKDSGNKPPGLL 59  
Db 47 PLPTSPMNLKMQSQAPPTSWLAKRLFCPPPHWK-----SLTASOKSFSKAAK 99  
QY 60 PRK-GLYMANDKLLRHHLQIPIHPKDFLSVMLEKGS-----SAMRELTAVNLEH 110  
Db 100 PRPGGSWMTFKRSSPSALGRANH-PQDRRLGYDGKGVRLASLDETQACNAFAIN-KA 157  
QY 111 PEMLE---KASRELWVRWNRNEDITEPOSILAAAKAG----- 146  
Db 158 PAILEGFVEFEV-----SVIAARDSGNVAIFDLAENVHDKGILATST 202  
QY 147 -----MSAEQAQGLLEKIA-----TPKVNOLKETTE 173  
Db 203 VPAAISVQTAEEAAAEKLLHALDYGVGLGFEFFVLKDTLLANEFAPRVHNS-GHWTE 261  
QY 174 AACRYGAF-----GLPI--TVAHVDGQTHMLFGSD-----RMELLAHLLEK 213

Db 262 AACAIQFEQHIRAVAGLPLGNTDRHSDCVMENLIGDDIEKVPAILCEKNAVLHLYGKK 320  
Search completed: February 15, 2001, 15:53:24  
Job time: 156 sec

